PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 18 January 15, 1932 Number 1

# THE GOMPERTZ CURVE AS A GROWTH CURVE

By Charles P. Winsor

Department of Biology, School of Hygiene and Public Health, Johns Hopkins University

Communicated December 2, 1931

In 1825 Benjamin Gompertz published a paper in the *Philosophical Transactions of the Royal Society*, "On the Nature of the Function Expressive of the Law of Human Mortality," in which he showed that "if the average exhaustions of a man's power to avoid death were such that at the end of equal infinitely small intervals of time, he lost equal portions of his remaining power to oppose destruction," then the number of survivors at any age $x$ would be given by the equation

$$L_x = kg^{c^x} \tag{1}$$

(It is clear that Gompertz means equal proportions, not equal absolute amounts, of the "power to oppose destruction.")

The Gompertz curve was for long of interest only to actuaries. More recently, however, it has been used by various authors as a growth curve, both for biological and for economic phenomena. It is the purpose of the present note to consider some of the mathematical properties of this curve, and to indicate to some extent its usefulness and its limitations as a growth curve.

For actual purposes the curve is generally written in the form (1); but for our purpose it is more convenient to write it

$$y = ke^{-e^{a-bx}} \tag{2}$$

in which $k$ and $b$ are essentially positive quantities.

From (2) it is clear that as $x$ becomes negatively infinite $y$ will approach zero, and as $x$ becomes positively infinite $y$ will approach $k$. Differentiating (2) we have

$$\frac{dy}{dx} = kbe^{a-bx}e^{-e^{a-bx}} = bye^{a-bx} \tag{3}$$

and it is apparent that the slope is always positive for finite values of $x$, and approaches zero for infinite values of $x$. Differentiating again we have

$$\frac{d^2y}{dx^2} = b^2ye^{a-bx}(e^{a-bx} - 1). \qquad (4)$$

From (4) we see that there will be a point of inflection when

$$x = \frac{a}{b}$$

The ordinate at the point of inflection is

$$y = \frac{k}{e}$$

or approximately, when 37 per cent of the final growth has been reached.

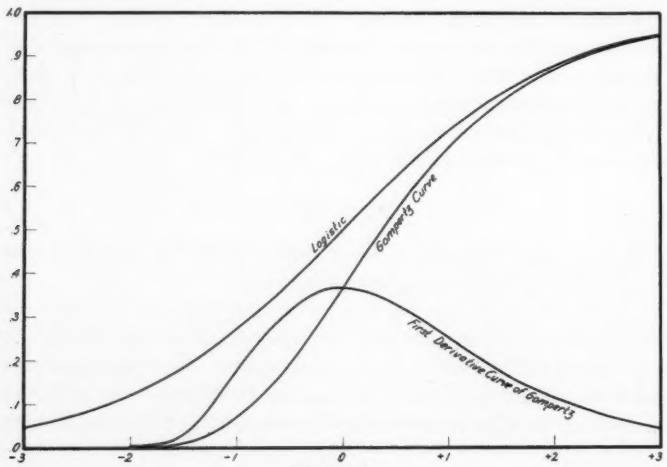



FIGURE 1

Figure 1 shows the form of the curve for the case $k = 1$, $a = 0$, $b = 1$; there are also shown the logistic and the first derivative of the Gompertz curve.

Equations:

Gompertz: $$y = e^{-e^{-x}}$$

Logistic: $$y = \frac{1}{1 + e^{-x}}$$

When we desire, therefore, to fit growth data which show a point of

inflection in the early part of the growth cycle, when approximately 35 to 40 per cent of the total growth has been realized, we may use the Gompertz curve with the expectation that the approximation to the data will be good. There seems, however, no particular reason to expect that the Gompertz curve will show any wider range of fitting power than any other three-constant *S*-shaped curve. For example, the logistic

$$y = \frac{k}{1 + e^{a-bx}} \tag{5}$$

possesses the same number of constants as the Gompertz curve, but has the point of inflection mid-way between the asymptotes. The degree of "skewness" in the Gompertz curve is just as fixed as in the logistic; and it is clear that to introduce a variable degree of skewness into a growth curve will require at least four constants.

To illustrate the mathematical properties of the Gompertz curve and the logistic, the following table has been prepared.

| PROPERTY | GOMPERTZ | LOGISTIC |
|---|---|---|
| Equation | $y = ke^{-e^{a-bx}}$ | $y = \frac{k}{1 + e^{a-bx}}$ |
| Number of constants | 3 | 3 |
| Asymptotes | $\begin{cases} y = 0 \\ y = k \end{cases}$ | $\begin{cases} y = 0 \\ y = k \end{cases}$ |
| Inflection | $\begin{cases} x = \frac{a}{b} \\ y = \frac{k}{e} \end{cases}$ | $\begin{cases} x = \frac{a}{b} \\ y = \frac{k}{2} \end{cases}$ |
| Straight line form of equation | $\log \log \frac{k}{y} = a - bx$ | $\log \frac{k - y}{y} = a - bx$ |
| Symmetry | Assymetrical | Symmetrical about inflection |
| Growth rate | $\frac{dy}{dx} = bye^{a-bx} = by \log \frac{k}{y}$ | $\frac{dy}{dx} = \frac{b}{k} y(k - y)$ |
| Maximum growth rate | $\frac{bk}{e}$ | $\frac{bk}{4}$ |
| Relative growth rate as function of time | $\frac{1}{y}\frac{dy}{dx} = be^{a-bx}$ | $\frac{1}{y}\frac{dy}{dx} = \frac{b}{1 + e^{-a+bx}}$ |
| Relative growth rate as function of size | $\frac{1}{y}\frac{dy}{dx} = b(\log k - \log y)$ | $\frac{1}{y}\frac{dy}{dx} = \frac{b}{k}(k - y)$ |

The parallelism between the Gompertz curve and the logistic may be carried further. It has been found useful, for example, to add a constant term to the logistic, giving it a lower asymptote different from zero

$$y = d + \frac{k}{1 + e^{a-bx}} \tag{6}$$

Clearly this procedure is equally applicable to the Gompertz curve, giving

$$y = d + ke^{-e^{a-bx}} \tag{7}$$

It is also clear that in general the sum, or the average, of several Gompertz curves will not be a Gompertz curve; just as several logistics do not, in theory, give a logistic when added or when averaged. But just as it has been found in practice that the sum of a number of logistics does in fact often approximate closely a logistic as has been shown by Reed and Pearl (1927), it will be true that Gompertz curves will often add to give something very close to a Gompertz curve. And the general theory of averaging growth curves worked out by Merrell (1931) and applied by her to the logistic can be applied without modification to the Gompertz curve.

It may be further pointed out that the Gompertz curve may be generalized in the manner in which Pearl and Reed (1923) have generalized the logistic.

Pearl and Reed set

$$\frac{dy}{dx} = \frac{b}{k} y(k - y)f(x) \tag{8}$$

and on integration obtain

$$y = \frac{k}{1 + e^{F(x)}} \tag{9}$$

and assuming that $F(x)$ can be expanded as a Taylor's series, they reach

$$y = \frac{k}{1 + e^{a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots}}. \tag{10}$$

In a similar fashion, the differential equation of the Gompertz curve may be written

$$\frac{dy}{dx} = by \,(\log k - \log y), \tag{11}$$

and if we add an arbitrary function of time on the right hand side of this equation

$$\frac{dy}{dx} = by \,(\log k - \log y)f(x) \tag{12}$$

we have on integration

$$y = ke^{-e^{F(x)}} \tag{13}$$

and if $F(x)$ can be expanded in a Taylor's series we have

$$y = ke^{-e^{a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots}} \tag{14}$$

It will be noted that if we wish to use only a finite number of terms in the power series, we must keep an odd power of $x$ for our highest term, if our curve is to run from $y = 0$ to $y = k$. This restriction applies to both the generalized logistic and the generalized Gompertz curve.

We may rationalize the derivation of the Gompertz curve along the lines indicated by Ludwig (1929). Ludwig postulates that the relative growth rate $\frac{1}{y}\frac{dy}{dx}$ must decrease monotonically with continued growth. If now we write

$$\frac{1}{y}\frac{dy}{dt} = m - ny \qquad (15)$$

we have the differential equation of the logistic. In this case the relative growth rate decreases as a linear function of growth already reached. If we set

$$\frac{1}{y}\frac{dy}{dx} = pe^{-qx} \qquad (16)$$

we have on integration the Gompertz curve. In this case the relative growth rate decreases exponentially with time—in Gompertz's phrasing, in equal small intervals of time the organism loses equal proportions of its power to grow.

This rationalization of the Gompertz curve is that used by Wright (1926), Davidson (1928) and Weymouth, McMillin and Rich (1931). (Ludwig does not consider the Gompertz curve.) Other authors (citations in literature list) who have used the Gompertz curve as a growth curve have not attempted a rational derivation, but have apparently been led to its use more or less empirically, much as Pearl and Reed (1920) were originally led to the logistic.

Wright (1926) appears to have first suggested the use of the Gompertz curve for biological growth. He says:

"In organisms, on the other hand, the damping off of growth depends more on internal changes in the cells themselves, the process which Minot called cytomorphosis. The average growth power as measured by the percentage rate of increase tends to fall at a more or less uniform percentage rate, leading to asymmetrical types of $S$-shaped curves of which the form $\log \log \frac{k}{y} = a(b - x)$ is a simple example, instead of the logistic curve $\log\left(\frac{k}{y} - 1\right) = a(b - x)$."

Following Wright, Davidson (1928) used the Gompertz curve to represent the growth in body weight of cattle. He pointed out that the ordinate

at the point of inflection is $\frac{1}{e}$ of the limiting ordinate. Davidson's figures actually cover only the upper part of the growth curve; his lowest weight is about 70 per cent of the limiting weight, so that no inflection appears in his actual data.

Weymouth, McMillin and Rich (1931) have used the Gompertz curve to represent the growth in shell size of the razor clam. They state that the curve also gives good fits for the guinea-pig and the rat. It must be noted that while their curves give good fits, they have found it necessary, in their most extensive series, to use two different curves to graduate the first and second halves of their data. They state:

"It is to be noticed, however, that after the eighth winter the slope is distinctly less and toward the end, due to inadequate data, the curve is irregular. It is true of all curves that, even before they become irregular because of small numbers, the curves deviate from the earlier trend, indicating that the growth is maintained at a higher rate than the first part of the curve would lead one to predict."

Weymouth and Thompson (1931) have also applied the Gompertz curve to the growth of the Pacific cockle.

It is of interest to compare this mode of deriving growth curves with that given by Lotka (1925). Lotka begins with the assumption that growth takes place under constant external conditions, so that the growth rate may be regarded as a function of attained growth alone:

$$\frac{dy}{dx} = F(y). \tag{17}$$

We now consider the equilibrium conditions, that is, the conditions where

$$\frac{dy}{dx} = F(y) = 0. \tag{18}$$

Clearly $F(y)$ must vanish for $y = 0$. Then if we assume that $F(y)$ can be expanded as a Taylor's series, the series will contain no constant term, and we shall have

$$\frac{dy}{dx} = ay + by^2 + cy^3 + \ldots \tag{19}$$

We now note that (18) must have at least one other root, if there is to be some upper limit to growth. The simplest case which satisfies this condition is where the series in (19) terminates with the term in $y^2$, so that

$$\frac{dy}{dx} = ay + by^2. \qquad (20)$$

But this is the differential equation of the logistic. It does not appear that this particular line of reasoning is likely to lead to the Gompertz curve.

*Summary.*—The Gompertz curve and the logistic possess similar properties which make them useful for the empirical representation of growth phenomena. It does not appear that either curve has any substantial advantage over the other in the range of phenomena which it will fit. Each curve has three arbitrary constants, which correspond essentially to the upper asymptote, the time origin, and the time unit or "rate constant." In each curve, the degree of skewness, as measured by the relation of the ordinate at the point of inflection to the distance between the asymptote, is fixed. It has been found in practice that the logistic gives good fits on material showing an inflection about midway between the asymptotes. No such extended experience with the Gompertz curve is as yet available, but it seems reasonable to expect that it will give good fits on material showing an inflection when about 37 per cent of the total growth has been completed. Generalizations of both curves are possible, but here again there appears to be no reason to expect any marked difference in the additional freedom provided.

I should like to express my thanks to Dr. Raymond Pearl, at whose suggestion this paper was written, for his advice and criticism.

## LITERATURE CITED

*A. Actuarial use of the Gompertz curve.*

Gompertz, Benjamin, "On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Method of Determining the Value of Life Contingencies," *Phil. Trans. Roy. Soc.*, 513–585 (1825).

King, George, *Institute of Actuaries Text Book*. Part II. "Life Contingencies." Second Edition. London. (Chapter VI, 68–104 (1902), contains an account of the Gompertz curve from an actuarial point of view.)

*B. Application to growth of organisms.*

Wright, Sewall, book review in *Jour. Am. Stat. Assoc.*, **21**, 494 (1926).

Davidson, F. A., "Growth and Senescence in Purebred Jersey Cows," *Univ. of Ill. Agr. Exp. Sta. Bull.*, No. **302**, 192–199 (1928).

Weymouth, F. W., H. C. McMillin, and W. H. Rich, "Latitude and Relative Growth in the Razor Clam, *Siliqua patula*," *Jour. Exp. Biol.*, **8**, 228–249 (1931).

Weymouth, F. W., and Thompson, Seton H., "The Age and Growth of the Pacific Cockle (*Cardium corbis*, Martyn)," *Bull. Bur. Fisheries*, **46**, 633–641 (1930–1931), Bur. Fish. Doc. No. 1101.

*C. Application to psychological growth.*

Courtis, Stuart A., "Maturation Units for the Measurements of Growth," *School and Society*, **30**, 683–690 (1929).

*D. Application to population growth.*

Davies, George R., "The Growth Curve," *Jour. Am. Stat. Assoc.*, **22**, 370–374 (1927).

Will, Harry S., "On Fitting Curves to Observational Series by the Method of Differences," *Ann. of Math. Stat.*, **1**, 159–190 (1930).

*E. Application to economic growth.*

Prescott, Raymond B., "Laws of Growth in Forecasting Demand," *Jour. Am. Stat. Assoc.*, **18**, 471–479 (1922).

Peabody, Leroy E., "Growth Curves and Railway Traffic," *Jour. Am. Stat. Assoc.*, **19**, 476–483 (1924).

*F. Other literature on growth curves.*

Pearl, R., and Reed, L. J., "On the Rate of Growth of the Population of the United States and Its Mathematical Representation," *Proc. Nat. Acad. Sci.*, **6**, 275–288 (1920).

Pearl, R., and Reed, L. J., "On the Mathematical Theory of Population Growth," *Metron*, **3**, 6–19 (1923).

Lotka, Alfred J., *Elements of Physical Biology*, Baltimore. Chap. VII, 64–66 (1925).

Reed, L. J., and Pearl, R., "On the Summation of Logistic Curves," *Jour. Roy. Stat. Soc.*, **90**, 729–746 (1927).

Ludwig, Wilhelm, "Vergleichende Untersuchung über Wachstumsgesetze," *Biol. Zentralbl.*, **49**, 735–758 (1929).

Merrell, Margaret, "The Relationship of Individual Growth to Average Growth," *Human Biology*, **3**, 37–70 (1930).

---

## *REVERSIBLE COAGULATION IN LIVING TISSUE. X**

BY WILDER D. BANCROFT, ROBERT S. GUTSELL,[1] AND JOHN E. RUTZLER, JR.[2]

BAKER CHEMICAL LABORATORY, CORNELL UNIVERSITY

Communicated December 21, 1931

The final test of the theory, that agglomeration accompanies addiction to morphine, is to show that a peptizing agent reduces the severity of withdrawal symptoms or else eliminates them entirely. Another phase of the problem is to show that the administration of a peptizing agent to an individual withdrawn from morphine stops post-withdrawal symptoms, particularly the craving for the drug. To this end a morphine addict was withdrawn from the drug while treating him with sodium rhodanate as the theory[3] demands.

J. H., male, aged 49, a trained nurse, presented himself for treatment by us. He had been addicted to the use of morphine for 16 years, during which time he was withdrawn from the drug about six times. The methods of treatment used upon him were: insensible withdrawal, slow withdrawal, abrupt withdrawal and the Towns-Lambert treatment. Once the patient did not use morphine for four months after being withdrawn; other times he resumed the use of the drug in from two days to three or four weeks.

At the time that this treatment was started the patient was using 25 grains of morphine in two days, injecting it intravenously. An additional

factor is a diagnosis of manic depressive depressed, which was made at a State Hospital about a year ago.

Upon admission to a private room in the Ithaca Memorial Hospital,[4] the patient's clothing and person were thoroughly searched for concealed morphine. At this time a hypodermic syringe and needle were taken from him and disposed of. A thorough physical examination showed that the general health of the patient was good; he was thin, but not emaciated.

The patient was put to bed upon admission to the hospital and kept there until three days after the morphine was stopped. During the whole course of the treatment the patient was not allowed to have visitors except members of the immediate family, and then only when accompanied by the nurse in attendance. He was not allowed to have any packages, and the two or three letters that came for him were carefully scrutinized before being given to him.

Throughout the course of treatment several laboratory tests were made at frequent intervals. Daily leucocytes and differential blood counts were made. The basal metabolic rate was determined several times. The blood pressure was recorded frequently. The specific gravity, volume and acidity of the 24-hour urine was determined often. At the same time tests were made for sugar and albumin in the urine. A microscopic examination was also made. Kidney function was determined periodically. Pulse rate, respiration rate and temperature were recorded routinely. The sedimentation rate of the red blood corpuscles and rouleau formation were followed closely. Psychological tests were made to determine fatigue, accuracy and concentration throughout withdrawal period. Controls on these tests were made by a two-day study of the patient before administering sodium rhodanate or reducing the dose of morphine.

Upon admission, and for two days afterward, the patient appeared to be calm with a definite substratum of apprehension and nervousness. Morphine was administered hypodermically rather than intravenously. The patient's skin was dry. He was bothered by frequent urinations and an inability to empty the bladder completely. It was definitely determined that this was not due to enlargement of the prostate gland.

TABLE I

| DATE | GRAINS MORPHINE | GRAMS NaCNS | DATE | GRAINS MORPHINE | GRAMS NaCNS |
|---|---|---|---|---|---|
| 12–1–31 | 6 | 0 | 12–6–31 | 2.5 | 1.33 |
| 12–2–31 | 12 | 0 | 12–7–31 | 1.0 | 1.10 |
| 12–3–31 | 7.5 | 0.67 | 12–8–31 | 0.62 | 3.00 |
| 12–4–31 | 6 | 0.67 | 12–9–31 | 0 | 3.8 |
| 12–5–31 | 4 | 1.33 | 12–10–31 | 0 | 1.5 |

When the preliminary laboratory study was completed we began to administer sodium rhodanate and reduce the dosage of morphine. Table I

shows how this was done. Thus, in six days the patient was reduced from 12 grains of morphine to none. The patient was given only six grains of morphine the first day because he was probably overloaded with the drug upon admission. That he was accustomed to more morphine than this was evidenced by the fact that he became very nervous on the day he had only six grains.

Table II shows the changes in blood pressure and weight during the course of the treatment.

TABLE II

| DATE | BLOOD PRESSURE | WEIGHT IN POUNDS | DATE | BLOOD PRESSURE | WEIGHT IN POUNDS |
|---|---|---|---|---|---|
| 12–1–31 | 115/70 | ... | 12–8–31 | 122/68 | 138 |
| 12–2–31 | 105/65 | 139 | 12–9–31 | 120/74 | 139 |
| 12–3–31 | 120/70 | 140 | 12–10–31 | 130/75 | 139 |
| 12–4–31 | 112/60 | 141 | 12–11–31 | 135/78 | ... |
| 12–5–31 | 115/70 | 141 | 12–12–31 | 135/74 | 135 |
| 12–6–31 | 113/68 | 142 | 12–13–31 | .... | 135 |
| 12–7–31 | 115/68 | 141 | | | |

Data show that there was a slight increase in the blood pressure during the treatment despite the fact that sodium rhodanate is reputed to lower the blood pressure The loss of weight was not due directly to the few withdrawal symptoms that the patient experienced, but rather to loose bowels. The patient had no morphine after the eighth of December and the slight loss in weight did not appear until three days later.

The daily study of the patient's blood revealed the fact that the usual leucocytosis[5] developed on the day of withdrawal; this is shown in Table III.

TABLE III

HEMATOLOGY[6]

| DATE | W.B.C. | POLYS. | LYMPS. | MONOS. | EOSINS. | BASOS. |
|---|---|---|---|---|---|---|
| 12–1–31 | 7,850 | 66% | 30.5% | 1.0% | 2.0% | 0.5% |
| 12–2–31 | 9,750 | 64.5% | 29.5% | 2.0% | 3.0% | 1 |
| 12–3–31 | 10,300 | 73.5% | 23.5% | 0.5% | 2.0% | 0.5% |
| 12–4–31 | 8,525 | 61.5% | 34.5% | 0.5% | 2.0% | 1.5% |
| 12–5–31 | 9,000 | 64.0% | 33.0% | 2.0% | 0 | 1.0% |
| 12–7–31 | 10,600 | 66.0% | 31.5% | 1.0% | 1.0% | 0.5% |
| 12–8–31 | 13,200 | 77.0% | 20.0% | 3.0% | 0 | 0 |
| 12–9–31 | 14,600 | 73.5% | 24.5% | 1.5% | 0 | 0.5% |
| 12–10–31 | 14,900 | 77.5% | 20.0% | 2.0% | 0 | 0.5% |
| 12–11–31 | 11,550 | 74.5% | 23.5% | 2.0% | 0 | 0 |
| 12–12–31 | 10,000 | 62.0% | 36.0% | 2.0% | 0 | 0 |
| 12–14–31 | 11,300 | 69.0% | 29.0% | 2.0% | 0 | 0 |

Further studies were made during convalescence; they show that the leucocytosis became less marked and then increased. We attribute this

second increase to the sodium rhodanate; for in normal rabbits this compound causes a leucocytosis.[7]

Table IV shows the urinalysis data. This table reveals nothing out of the ordinary except the expected increase in specific gravity of the urine, which is undoubtedly due to presence of sodium rhodanate and an increased amount of sodium chloride. The sodium chloride is released from the tissues because of the stronger adsorption of the sodium rhodanate.

TABLE IV

| DATE | APPEARANCE | REACTION | SP. GRAV. | ALBUMIN | SUGAR |
|---|---|---|---|---|---|
| 12–1–31 | Cloudy amber | Alkaline | 1,025 | Trace | ... |
| 12–3–31 | Clear straw | Acid | 1,021 | " | ... |
| 12–4–31 | Cloudy straw | Acid | 1,021 | " | ... |
| 12–5–31 | Cloudy straw | Acid | 1,016 | " | ... |
| 12–6–31 | Cloudy amber | Acid | 1,025 | ... | ... |
| 12–7–31 | Cloudy amber | Acid | 1,027 | Trace | ... |
| 12–8–31 | Cloudy amber | Alkaline | 1,019 | " | ... |
| 12–9–31 | Cloudy amber | Alkaline | 1,031 | " | ... |
| 12–10–31 | Cloudy amber | Alkaline | 1,032 | " | ... |
| 12–12–31 | Cloudy amber | Acid | 1,035 | " | ... |

Metabolism tests were taken at different intervals and since plus or minus 10 is normal, the patient's tests were all within normal limits. The rates are given in Table V.

TABLE V

| DATE | RATE |
|---|---|
| 12–2–31 | + 1% |
| 12–5–31 | −10% |
| 12–8–31 | − 1% |
| 12–12–31 | + 8% |
| 12–19–31 | + 3% |

Kidney function tests were made which show that there was a normal permeability to phenolsulphonphthalein, since 90 to 93% of the dye was generally eliminated within two hours after being injected intravenously. From this we can safely conclude that sodium rhodanate in comparatively very large doses has no serious irritating effect on the kidneys.

Sedimentation velocity tests on the red blood corpuscles were made; but due to there being at least two factors affecting the rate of sedimentation, namely, morphine and sodium rhodanate, conclusions cannot be drawn without further study. However, it was demonstrated definitely that when sodium rhodanate is added to the blood, which is taken for the sedimentation test, the amount of sedimentation in a given time is very markedly reduced.

The most valuable observations made on this case were those of a clinical

nature, involving as they do the symptomology of withdrawal. During the second day that the patient was under observation, he was given 12 grains of morphine; as a result he was more quiet than on the preceding day. Otherwise there was no abnormal reaction to this dosage. This leads one to believe that 12 grains are about the quantity of morphine that the patient was used to.

On 7.5 grains of morphine and 10 grains of sodium rhodanate during the third day, the patient was quiet, friendly, coöperative, ate well and slept well. At times he would try to convince those attending him that, although he was a drug addict, his family was respectable. The manic depressive characteristics of the patient were beginning to come to the surface. He was argumentative and very talkative. It became very obvious that the patient was extremely jealous of his own personal comfort.

The fourth day saw an improvement in the patient's appetite. During much of the day he was greatly relaxed and neither nervous nor apprehensive. He remarked that some of the wrinkles had gone out of his face. He slept off and on during the afternoon, and yawned a great deal. The patient received six grains of morphine and 1.5 grams of sodium rhodanate during the day. His evening dose of sodium rhodanate had a definite quieting effect.

On the fifth day the patient was more composed, somewhat less talkative and apparently normal, mentally and physically. This day was the best that he had had up to this time. There was no noticeable substratum of apprehension. During these five days the patient had slept more each night. These conditions prevailed on a total dose of only four grains of morphine.

On the sixth day we saw a slight change in the condition of the patient. At noon time he complained of cramps which were caused by the action of a physic on a constipated bowel, there having been no evacuation for three days. In the afternoon the patient appeared to be somewhat more nervous. He stated that when he is withdrawn from morphine, he can tell to within $^1/_8$ grain how much of the drug he is getting. Shortly thereafter he was asked to estimate the quantity of morphine that was being given to him in each dose. He was totally unable to estimate the quantity. Daily from this time on the patient was asked to guess how much morphine was being administered in each dose. His estimation invariably was 100% or more too high. The patient was not told whether his guesses were right or wrong, so that he did not know how much of the drug he was getting until the last reduction in dosage was made.

During the afternoon of the fifth day, the patient had spells during which he was unable to concentrate his attention well. Each time that

such a condition was observed, the patient became markedly more excited a short time afterward. The action of the sphincter urethrae was normal from this time on. The patient voided easily in a good stream and emptied his bladder.

During the evening of the sixth day, the patient was very nervous, slightly depressed, perspired a little, could not stay in bed for long spells and his muscles and back ached slightly. These were all the symptoms that were observed on a total dosage of morphine of only $2^1/_2$ grains.

Many times during this treatment it was observed that the sodium rhodanate quieted the patient down without the administration of morphine.

The seventh day was essentially the same as the sixth, until evening, when continuous motor activities were observed while the patient lay on the bed, mentally calm and collected. This motor activity may be due to a slightly increased concentration of morphine in the cord and the decrease of the drug in the brain and sensory nerves, due to the action of the rhodanate ion. For the first time during the course of the treatment, the pupils of the eyes were normal in size. At no time before or after this was there any dilatation of the pupils. The dosage of morphine was 1 grain on this day.

Early in the morning of the eighth day the patient was nervous, perspired a little, had a slight backache and was slightly depressed. There was a brief wave of discouragement and severe nausea. As the day progressed, he became more nervous, so that finally he could not stay in bed or stay in a chair or walk around the room for more than five minutes at a time. This was the only withdrawal symptom that was marked enough to be considered as definitely a part of the picture. In the afternoon he was given 6 grains of sodium amytal in capsules by mouth. He took it sullenly. A half-hour later the patient was more excited, restless and nervous. At this time 3 grains more of sodium amytal were given. When another half-hour had elapsed, the patient was thrashing about on the bed. He was irrational, definitely delirious, had no sense of balance, had lost his memory, was ataxic, confused, intoxicated and showed hypermotor activity. We later learned that during this time he was not suffering from withdrawal symptoms because almost the whole of the experience was a blank in his memory. An injection of morphine cleared him mentally to a certain extent. From this time on he became more clear and less restless, although the process was slow.

Late in the evening of the eighth day, the patient was told that he was getting only an eighth of a grain of morphine at a dose. With a wide grin on his face he turned over in bed and said that we would not hear another word out of him. He quieted down immediately and asked for food; this was the first that he had had in twelve hours.

The fact that the patient quieted down very markedly immediately upon being told that he was getting only one-eighth of a grain of morphine at a dose indicates quite clearly that most of the withdrawal symptoms that were manifested were due to the fear that he had many days of suffering before the morphine was entirely withdrawn.

The early hours of the morning of the ninth day found the patient restless and talkative. He perspired slightly. During the day the patient's face flushed up several times. The reaction did not last long; and his face did not at any time become unusually white. There were definite auditory hallucinations, and at times mental discontinuities. These were especially evident when the patient was on the verge of sleep. His appetite improved. During the evening the patient quieted down considerably; he was not depressed at all. Throughout this study there was no serious depression observed. After the ninth day there were no signs of depression at all. The patient was incredulous when he was told during the evening of the ninth day that he had had no morphine for 24 hours.

On the tenth day the patient was completely relaxed and comfortable. There were no signs of withdrawal symptoms whatever, except that he had six loose bowel movements. Even the bowels became normal after the tenth day.

On the eleventh day the patient was up and dressed; during the twelfth day the patient vomited several times, usually bringing up only bile. That this was due to the action of the sodium rhodanate seems evident, for it stopped when the dose was changed and given after meals.

About the thirteenth day the patient began to complain bitterly of periods of sleeplessness at night. This difficulty was circumvented by giving large doses of sodium rhodanate just before the patient retired.

Data obtained from the physician in charge during a previous withdrawal elicited the information that he was a "mean customer," that he had periods of severe depression; that he suffered a severe gastro-intestinal upset; this was characterized by cramps, diarrhea and nausea. There was profuse sweating, and a marked sensitivity to heat and cold. He suffered severely all the common withdrawal symptoms. On withdrawal under sodium rhodanate none of these effects were observed. Likewise, of the multitude of withdrawal symptoms that this addict has been through under previous treatments, and which most addicts manifest, none but those few which have been mentioned have been observed with this treatment. There were no post-withdrawal symptoms whatever that were not completely controlled by the administration of sodium rhodanate.

Since the patient has been withdrawn he claims to have had no desire whatever for morphine, and our observations of him support his contention.

In so far as this case is typical of the response of drug addicts to with-

drawal when treated with sodium rhodanate the reversible agglomerating action theory of the action of morphine is fully substantiated.

The conclusions to be drawn from this paper are:

1. In one case morphine has been withdrawn from a drug addict of long standing without the appearance of any major withdrawal symptoms.

2. The withdrawal symptoms were counteracted by administering sodium rhodanate to the patient in fairly large quantities while reducing the dosage of morphine.

3. The action of the sodium rhodanate is to peptize rapidly the protein colloids that are agglomerated by morphine, thus passing very quickly through the excitement stage, during which there are withdrawal symptoms.

4. Sodium rhodanate probably hastens the elimination of morphine from the addicted tissue by displacing the adsorbed morphine.

5. The results obtained by us when working with dogs have been substantiated in the case of a human being.

6. It has been demonstrated that the craving for morphine can be controlled by sodium rhodanate, which means that mental rehabilitation will almost surely follow the treatment, by the use of this peptizing agent, of those who have been withdrawn from morphine.

We understand unofficially that good results have already been obtained on drug addicts by other physicians who have been testing this theory; but, unfortunately, nothing has been published so far and we have no first-hand information as to the results.

* This work is done under the programme now being carried out at Cornell University and supported in part by a grant from the Heckscher Foundation for the Advancement of Research established by August Heckscher at Cornell University.

[1] M.D.

[2] Eli Lilly Research Fellow.

[3] Bancroft and Rutzler, *Proc. Nat. Acad. Sci.*, **17**, 186 (1931).

[4] The authors wish to thank those connected with the Ithaca Memorial Hospital for their unfailing courtesy and readiness to help at all times during the progress of this study.

[5] Lambert, *Am. J. Psychiatry*, **10**, 433 (1930).

[6] Dr. Hauenstein and his staff of the Tompkins County Laboratory have always been ready to help with the studies that we have made on reversible agglomeration in living tissue; this time has not been an exception.

[7] Takacs, *Z. ges. exptl. Med.*, **50**, 432 (1926).

# THE ARIZONA EXPEDITION FOR THE STUDY OF METEORS

BY HARLOW SHAPLEY, ERNST J. ÖPIK, AND SAMUEL L. BOOTHROYD

HARVARD COLLEGE OBSERVATORY

Read before the Academy Tuesday, November 17, 1931

The nature of the interstellar and intergalactic media through which radiation, stars, clusters, and galaxies move is found to be of so much significance in our understanding of galactic distances and structure that fundamental research on the contents of space has become necessary. For several years at the Harvard Observatory we have studied one aspect of the problem—the meteors. Investigation of these multitudinous small bodies directly bears not only on knowledge of their own physical nature and their place in the cosmic structure, but as well on the question of the content of interstellar space; and indirectly such investigations may contribute to the solution of the general problem of "planetesimals" in the origin of the solar system, and of the structure of the upper terrestrial atmosphere.

Notwithstanding the superiority of the photographic plate over the human eye in most astronomical studies, in the field of meteoric astronomy photography still plays a minor rôle. Several hundred meteors are visible to the unaided eye to one that can be photographed. The Harvard collection of plates is, nevertheless, systematically searched for photographed meteor trails. The plates have occasionally yielded results unobtainable by ordinary visual methods, though they have contributed little toward the solution of the more important meteor problems, and have scarcely touched the field of such objects fainter than magnitude zero.

Among the investigations based on the Harvard photographs, some of which have been carried out with the aid of grants from the J. Lawrence Smith Fund of the NATIONAL ACADEMY of SCIENCES, we should mention: (*a*) distribution of observed fireballs throughout the year, showing principal maxima in August, November, and December, at the times of the Perseids, the Leonids and the Andromedes, and the Geminids; (*b*) the curvature of the paths of meteors in the earth's atmosphere; (*c*) the phenomena of multiple trails; (*d*) the occurrence, spacing, and photometry of "spindles" on the photographic trails; (*e*) the heights of meteors photographed at more than one station; (*f*) the paths of conspicuous fireballs as recorded by many observers.

During the past year the problem of the visual observation of meteors has been taken up in a systematic way. Dr. Öpik of Tartu, Estonia, joined the staff in 1930 in order to assist in organizing the observing

program and to take the leading part both in supervising a special expedition for the study of meteors and in the discussion of the results.

During the year, also, special grants were obtained from the Milton Fund of Harvard University, and from the Rumford Fund of the American Academy of Arts and Sciences for preliminary experimental work. Larger grants from the Wyeth Bequest to Harvard University and from the Rockefeller Foundation made the expedition possible. Professor Boothroyd of Cornell University joined the expedition, and a grant from the Heckscher Fund of Cornell was obtained to cover a part of the expenses.

In March, 1930, Dr. Shapley visited sites in Arizona, and chose Flag-

One of the two meteor houses used for visual observations in Arizona. In the window is seen the reticule giving coördinates for positions of meteors.

staff as headquarters. Dr. V. M. Slipher, director of the Lowell Observatory, very generously offered the coöperation of his staff and facilities of the observatory for the meteor observers and also permitted one of the meteor stations to be set up on Mars Hill near the buildings of the Lowell Observatory.

During the spring and summer of 1930 various types of experimental equipment were built and tested at the Harvard Observatory. The reticule house, shown in the accompanying photograph, was built in Cambridge, and after being tested was shipped to Flagstaff, where a duplicate house was made. Dr. Öpik and Professor Boothroyd, each with one

assistant, went to Arizona in the latter part of September, and systematic observations of meteors were begun early in October. After the program was inaugurated the equipment was supplemented in many details, and the full plan of observation was finally under way in November.

The details of the equipment, methods of observation, and problems under study are described briefly in the following paragraphs. In advance, we may say that the results promise to be even better than we had optimistically hoped. The observing conditions in the vicinity of Flagtaff appear to be excellent for researches on meteors; the number of meteors recorded is large, and the routine observers show the necessary skill and sustained enthusiasm.

The problems to be attacked by the expedition may be briefly summarized as the study of the cosmical origin of meteors. What is the distribution of their heliocentric velocities, both with regard to size and direction; what is the relative proportion of solar and hyperbolic meteors; the distribution of velocities of the hyperbolic meteors outside the solar system; the frequency-function of the relative masses of meteors; the absolute density of meteoric matter in space, in and outside the solar system? These are the questions to which answers in general outline may be expected from one year's observations by six observers under the favorable sky of the Arizona plateau.

At present the plan of observation is as follows: Several visual observers, working at two stations placed at a suitable east-west distance from each other, register meteors in two areas of about sixty degrees effective diameter centered on the meridian at forty-five degrees zenith distance north and south. To maintain statistical homogeneity of the records the scheme of fixed zenith distance and azimuth of the areas of observation was adopted. Iron reticules projected on the sky are used as coördinates of reference for tracing the meteor trails. These reticules represent, at ten degree intervals, declination and hour angle, except for the polar hour angles; the hour angle is easily transformed into right ascension when the time of observation is given. The reticules are mounted on two opposite slopes of a roof of a wooden shelter, which provides protection from wind and cold for the two observers working inside. The observer looks through one eye-hole of 32 mm. diameter placed at a distance of 50 cm. from the center of the reticule. In surveying the area, the observer has to move only his eye; the systematic error of parallax introduced in this way is small (a fraction of a degree only) and may be taken into account in the reductions. Without any corrections the reticules permit direct readings of celestial coördinates with an error not exceeding half a degree; with systematic corrections, the method is correct to within one tenth of a degree, and is thus much more accurate than individual meteor observations.

The widths of the wires, 0°.5 and 0°.8, permit them to be seen easily on the sky without artificial illumination; ten per cent of the area is obstructed by these wires, but the consequent loss in meteor numbers cannot exceed two per cent, because of the considerable length of the meteor trails as compared with the width of the wires. The records consist in tracing the trails on maps giving exact reproductions of the reticules, including thickness of the wires, on the scale 1:4. The time of appearance, to the nearest second (ascertained with the aid of a stop-watch), the magnitude, and the duration are also recorded.

The reticules cover 80° × 80° square (central dimensions), of which the observer can effectively watch a field 60° in diameter. The observers at the two stations watching the same region center their attention on areas displaced by the average amount of the parallactic displacement which is most conveniently chosen equal to twenty degrees; in this way the number of meteors observed in common is the greatest; at the same time, the average parallactic displacement being small as compared with the diameter of the field, the chances for meteors of unusually great or small height, seen by one observer, to fall out of the region of the other observer are not large (thirty per cent at the most). About forty kilometers distance between the two stations will give the required average displacement of twenty degrees at forty-five degrees zenith distance.

The visual reticule observations as described here are expected to give the relative meteor frequency for different hours of the night and different seasons of the year, and the frequency of apparent directions in the two areas (north and south) at different hours and seasons. With the aid of methods analogous to those used in the analysis of stellar proper motions, though more complicated, the frequency of directions will give the geocentric distribution of the directions of meteor trajectories, or the geocentric density of radiant points on the celestial sphere, the result being free of any hypothetical assumptions. Data on a great number of individual heights and real paths will be obtained; ample material, founded on a strict statistical basis, for the study of *group radiants** of meteors will be available, such a study being actually necessary in the course of the general analysis of the results.

Simultaneous and independent observations furnish a good means for studying observational selection depending upon brightness, duration, direction of motion, position in the area, and personality. The final statistical data must be corrected for the effect of selection. The reticule observations are supplemented by visual observations of meteor velocities, and by telescopic observations.

For the observation of angular velocities a device is used which may be called the "double-pendulum apparatus." It consists of a six-inch square plate glass mirror, resting freely on three supports forming an isosceles

right triangle; the support at the right angle is motionless, while the two other supports are tilted in the vertical direction, making harmonical oscillations with a difference of phase of ninety degrees. In this way the normal to the mirror plane is given a conical motion. In the case of equality of amplitudes of the two oscillations, a reflected stellar image will describe a circle in the zenith and an ellipse outside the zenith, the ratio of the axes being equal to Cos Z (assuming small amplitude), and the major axis being vertical. A meteor will describe a more complicated curve produced by the superposition of the elliptical oscillation upon its own motion. The shape of the apparent trajectory will depend upon the dimensions of the ellipse of oscillation, its orientation with respect to the direction of motion of the meteor, the period of oscillation, and the angular speed of the meteor. Pseudo-cycloidal trajectories are formed, with closed loops in the case of slow meteors, and open for those of high speed. The angular speed may be derived, from the tracings of the trajectories on a map, in two independent ways: from the shape of the trajectory, and from the length of one complete oscillation, which, divided by the period, gives the speed.

In the present experiments, the period was adopted equal to one tenth of a second; both the component oscillations are transmitted, by a system of levers, from the same end of a small eccentric shaft, the difference in phase being attained by placing the two levers at right angles to each other. The major axis of the oscillation is about half a degree. The supports of the mirror rest on steel balls in a special framing, giving them freedom to roll a few tenths of a millimeter in a horizontal direction. The levers, under permanent tension supplied by the weight of the mirror, which exceeds several times the inertia of oscillation, are continuously pulling. Under such circumstances, the oscillations are without any dead space, and secondary vibrations are negligible. For power, an electric synchronous motor is used, working at sixty revolutions per second (frequency of alternating current), with a gear reducing in the ratio 1:6.

The ellipses of oscillation in the case of stars can be seen as faint as the fourth magnitude, and at an angle not more than five degrees from the line of direct vision, their observation being thus limited to a field of only ten degrees effective diameter. For meteors, however, the circumstances are much more favorable, because the linear dimensions of the oscillation are increased along the trail and the resolving power of the eye is thus greater; also, objects of perceptible duration allow the observer to turn his eye upon them before the phenomenon ends, a condition favoring the selection of slow objects in velocity observations. From the experience of the first month of work with this apparatus, it appears that about eighty per cent of the number of meteors seen by a reticule observer may be observed and traced on a map. For about half of these, more or less

complete velocity criteria can be obtained. The fifth magnitude and an effective area of thirty degrees diameter represent approximately the limit of velocity observations with the present apparatus.

The systematic errors of observation may be checked in two different ways, namely, by meteors belonging to certain well-studied showers (in which case the effect of resistance of the air is a hypothetical factor, though not of great importance), and by laboratory experiments. Generally, the observations are not so difficult as they might seem at first, and with some experience the observer can record, in addition to the velocity criteria, the elements of position of the meteor with the same accuracy as in direct observations. The tracings are made on mirror image maps of the circumpolar area; stars are used for reference. In addition, a large scale drawing of the shape of the observed trajectory is made in the note-book.

In combination with the results of the reticule observations, the velocity data will aid in settling the fundamental problem of the cosmical origin of meteors; the distribution of geocentric radiant points deduced from the reticule data, together with the distribution of geocentric velocities furnished by the double pendulum observations, will permit derivation of the distribution of heliocentric velocities of the meteors, which is equivalent to the solution of the problem.

The telescopic observations not only form a necessary supplement to the visual ones, but are of independent value. There are indications that the statistical constitution may be different for meteors of different luminosity, and it becomes highly important to study a group that differs widely in luminosity from the visual meteors.

Simultaneous observations made by two 4-inch telescopes at two stations separated by about three kilometers distance in an east-west line form a substantial part of the program. The telescopes have a field of slightly less than 4 degrees, with eye-pieces of magnification seventeen times. As coördinates of reference, rectangular reticules in the focal planes are used, their exact positions being checked every night by transit observations of certain stars. The centers of the fields are in forty-five degrees zenith distance and are displaced by about $1^{\circ}.4$ with respect to each other in the direction of the average expected parallactic displacement, for the same reason as that given in the case of the visual observations. The observations are made north and south of the zenith on alternate nights.

After the first experiments with the visual velocity apparatus were made, it was decided to arrange velocity observations also with one of the two telescopes. The advantage, when compared with the visual apparatus, is that with the greater resolving power a greater number (practically all meteors seen) can be observed for velocity, and the accuracy must be much higher. Dr. C. O. Lampland, of Lowell Observatory, devised an

apparatus, which gives the same effect, on a different technical principle, as the double-pendulum apparatus. It involves an elliptical oscillation of the image without rotation of the mirror itself, the necessary conical movement being directly transmitted to a bar, normal to the mirror plane and rigidly connected with it. A plate glass mirror gave surprisingly good images with, of course, low magnifying power. The revolutions number thirty per second, the amplitude is adjustable and provisionally fixed at about $0^{\circ}.13$. The observations consist in tracing the apparent trajectory on a map which is a copy of the focal reticule.

The results of these telescopic observations will be similar to the combined visual results, but will yield a smaller amount of material; heights also will be available, but no real paths because of the small absolute size of the parallactic displacement (radiant point, or intersection of the trails observed at the two stations, becomes indeterminate). The real paths determined from visual observations, moreover, will be of statistical rather than of individual value, because for the predominantly short meteors the observational errors in direction are still considerable compared with the parallactic displacement of about twenty degrees.

The station on Mars Hill consists of one reticule house and the velocity apparatus, both visual and telescopic; the second visual station is twenty-three miles to the west, and consists only of the second reticule house. The second telescope, without velocity apparatus, is mounted two miles east of the Lowell Observatory.

Nights with bright moonlight are omitted from the program. During one lunation, observations are carried on for twenty-three days, with from four to eight hours of observation on one night. Of the 150 hours of possible observation per lunation, about eighty-three per cent are utilizable. When the hours of darkness exceed eight per night, observations are divided into two equal sessions, with an interval of one and a half to two and a half hours for sleep at midnight; such a program is easier to bear than six uninterrupted hours. Regularity of sleep and rest is of the utmost importance in securing homogeneous observations through all hours of the night.

The observers were carefully chosen after trial, especial care being taken to find persons interested scientifically in the matter. In October the staff of the expedition was composed, in addition to two of the authors of this communication, of Mr. George Mussen, of Cornell University; Mr. Helmut Haendler, of Harvard; Mr. Roger Wilson, of California Institute of Technology; Mr. Donald Hargrave, of Flagstaff. At the end of October Mr. Haendler left and was replaced by Mr. Robert Harris, of the University of New Mexico. In November, Mr. George Peters, of the University of Arizona, joined the staff.

The results of the first month of the incomplete program ending

November 6th are as follows: The total number of records amounts to 2755, covering about 2300 individual meteors; about 400 were observed simultaneously at the two stations, the data giving their heights and real paths; 234 velocity records with the double-pendulum apparatus were registered, of which about 70 indicate real paths. From these data one may estimate what the total number of records will be after the program of one year is completed (with the addition of about thirty per cent for telescopic observations).

* The term "group radiant" corresponds to "radiants" in the ordinary sense of meteor astronomy, i.e., several objects observed at nearly the same time having similar direction of motion and possibly being physically connected, directly or by origin. "Radiant point" generally will denote here a point on the sphere determining the direction of motion of a single body.

---

## *THE GROWTH OF SCENEDESMUS ACUTUS**

BY HOWARD SPRAGUE REED

UNIVERSITY OF CALIFORNIA CITRUS EXPERIMENT STATION, RIVERSIDE, CALIFORNIA

Communicated December 8, 1931

In organisms of a primitive type, reproduction and increase in number are substantially identical. A certain accretion in the mass of a one-celled organism is followed by the division of the nuclear and cytoplasmic substances into two daughter cells, each of which is morphologically similar to the cell from which it arose. In multicellular organisms which are differentiated into tissues and organs, the processes of reproduction and increase in cell number are by no means identical. In the first case the cell is continually losing its identity as a result of its division to form daughter cells; in the latter case the cellular masses soon differentiate into forms which retain their identity during the existence of the plant. The power of cell division is limited to a relatively small part of the multicellular organism. The increase in size of a multicellular plant is due in part to cell division, but mainly to cell enlargement.

The nature of the growth process in the two types of plants is a problem of interest in view of what has been briefly and inadequately outlined in the foregoing sentences. The growth curves of several annual and perennial plants have been studied and described in former years. The growth of Helianthus[1] and of shoots of Pyrus (Bartlett pear)[2] followed closely the curve of autocatalysis. In these cases the rate of growth of the plant was expressed by the equation $\frac{dx}{dt} = kx(A - x)$ in which $x$ = the size attained at any given time, $A$ = the final or limiting size, and $K$ = a constant.

In the cases cited, the equation represented the transformations and synthesis of materials into the substance of the plant. The mass of the plant depended upon the amount of transformable material in the system and the ability of the organism to transform the material.

Growth in size of cells is closely connected with the hydration capacity of protoplasm and other colloids which form important elements of the cell. Increase in number of cells is due to some factor about which we have little exact information. At present we have some evidence of a growth catalyst which directs the important activities which center about the cell nucleus, especially those activities concerned with the synthesis of new material.

The present investigation was planned in order to study the growth relationships of a unicellular alga (*Scenedesmus acutus*) with particular reference to the initial and final portions of the growth curve. As a problem in the growth of a population, this organism afforded an opportunity to study the development of a species of plants relatively simple in organization and without competition with other species in a closed system. Interest in this study lies in the fact that in a closed system the growth in numbers of organisms which have no organic dependence upon each other follows a curve of much the same type as the growth of multicellular organisms where enlargement is due both to cell multiplication and to cell enlargement.

The size of individual Scenedesmus cells is relatively constant. The number of individuals introduced at the outset for the purpose of inoculation was negligible in comparison with the volume of the nutrient solution employed. Except for the absorption of atmospheric carbon dioxide and oxygen, we had what may be termed a closed system. In this respect there is a profound contrast with the conditions under which one of the angiosperms lives and grows.

The observations on the growth of *Scenedesmus acutus* in pure culture were made while I was a guest in the Institute of Botany in the University of Geneva. It is a pleasure to express my gratitude to Professor Robert Chodat, director of the institute, for the many courtesies which I enjoyed while in his laboratory in the summer of 1930.

The alga used in these investigations grew in flasks containing sterilized Detmer's solution having an initial concentration of 585 parts per million. *Scenedesmus acutus* is especially suitable for cultures of this kind and for determination of the number of organisms per unit volume. It does not adhere tenaciously to the walls of the flask nor form large aggregates. The cells are generally separate, though aggregations of 4 and 8 cells are sometimes abundant. The culture flasks were kept at a temperature ranging from 20°C. to 24°C. in the diffused light of the laboratory. The average number of cells in the culture fluid was determined at intervals of 3 days or less.

A drop of well-mixed culture liquid was placed on a Thoma counting cell having a capacity of $^1/_{400}$ cu. mm., and the number of cells counted. From 18 to 35 samples were counted each time, the number depending upon the dispersion of the values about the mean. Densely populated samples were diluted with sterilized tap water before they were counted.

The population reached maximum density 5 or 6 weeks after inoculation, at which time the average number was about 4000 in the Thoma cells, or about 1,800,000 per cu. mm.

The numbers of organisms, when plotted against time (Fig. 1), give a

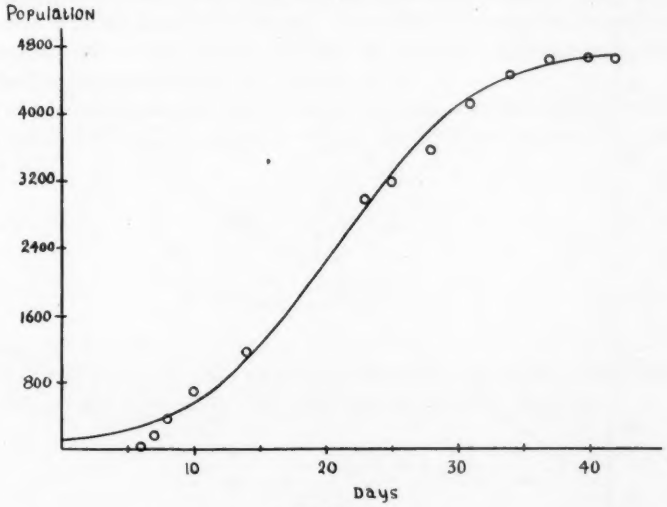



FIGURE 1

Growth curve of *S. aculus* calculated from the equation $\log \frac{x}{4450 - x} = 0.083(t - 20.6)$. Small circles represent average number of individuals observed on the days indicated.

close approximation to the sigmoid curve of autocatalysis derived from the equation

$$\log \frac{x}{A - x} = K(t - t_1). \qquad (1)$$

where $x$ = number of organisms at time $t$, $A$ = final or limiting number, $t_1$ = time at which $x = \frac{A}{2}$ and $K$ is a constant = $Ak$ of the differential equation. The population growth at the start lags far behind the calculated numbers, but after the first week there is fair agreement between the

observed and calculated values. The upper part of the curve approximates the observed values more closely. The initial lag period corresponds to that which I have previously observed in the case of shoots of trees and which Robertson observed in colonies of Infusoria. As the results stand, the probable error of a single observation is 89.4. The root-mean-square deviation is 127.4, or 5.09 per cent of the mean values of $x$. This is larger than the error of measuring or weighing larger organisms, but it is perhaps not abnormal for this type of observation.

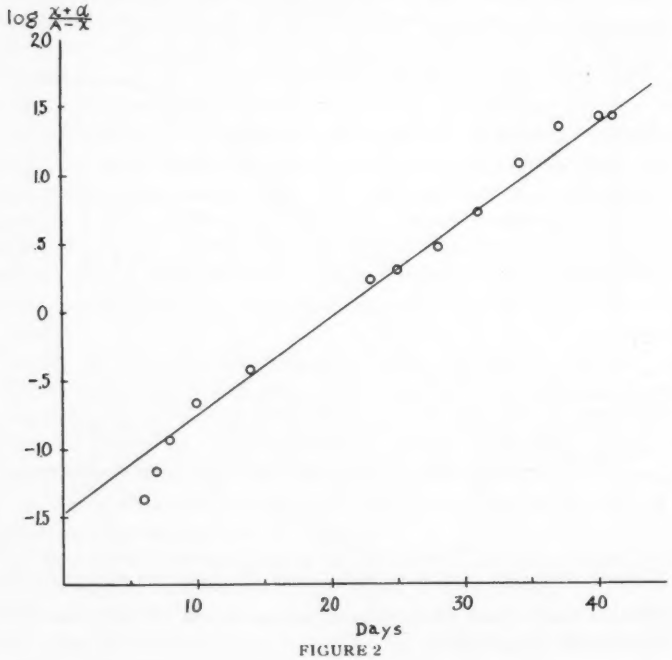



FIGURE 2

Evaluation of the constants of the equation $\log \frac{x+\alpha}{A-x} = \log \frac{\alpha}{A} + Kt$.

Important additions to our knowledge of autocatalysis have been made recently by Bray. A recent paper by Bray and Davis[3] points out the fact that a "pure" autocatalytic reaction will never start unless $x_1$ has a finite value.

The equation

$$\log \frac{x+\alpha}{A-x} = \log \frac{\alpha}{A} + t\frac{Ak}{2.303} \qquad (2)$$

evolved by Bray and Davis has been applied to the growth of Scenedesmus[4] with a somewhat greater degree of accuracy than the equation formerly used. In the case of a culture of unicellular organisms like that under consideration, $\alpha$ may be considered as the value of $x$ at zero time. In the case of multicellular plants it may represent the weight of the embryonic plant in the seed, or the height of the seedling when measurements begin. The values of $\alpha$ are nearly negligible in comparison with $x$ except in the early stages of growth.

The values of the constants were obtained by the following approximations and were finally tested graphically against the observed values. The value of $\alpha$ in the equation evidently represents the deficit in the number of organisms due to the action of some factor whose influence is especially evident in the early part of the growth curve.

The growth of the population seemed to be almost entirely inhibited for the first 3 days. If we regard $x = 0$ on the third day we may write equation (1)

$$\log \frac{x + \alpha}{A - x} = K(t - t_1)$$

$$\text{as } \log \frac{\alpha}{4450} = (0.083)\ (-17.6)$$

whence, $\alpha = 154$.

Since values of $x + \alpha$ are larger, it follows that values of $A$ will be larger, so we may take $A = 4550$. We may then write (2) in the form

$$\log \frac{x + \alpha}{A - x} - \log \frac{\alpha}{A} = t \frac{Ak}{2.303}$$

and determine the value of $t \frac{Ak}{2.303}$ for each value of $x$. The values of $\frac{Ak}{2.303}$ are then obtained by dividing each by $t$ and from these the mean value of $\frac{Ak}{2.303}$ was obtained. It was 0.071. For convenience we may write

$$\frac{Ak}{2.303} = K = 0.071.$$

Equation (2) then becomes

$$\log \frac{x + 154}{4550 - x} = \log \frac{154}{4550} + 0.071t \qquad (3)$$

and the results so obtained are given in table 1. The equation was tested by plotting values of $\log \frac{x + \alpha}{A - x}$ against $t$ (Fig. 2). The intercept on the

ordinate is $-1.47 = \log \frac{\alpha}{A}$ and the slope is $0.071 = K$. The agreement between observed and calculated values is good except in the first two cases.

TABLE 1

GROWTH OF SCENEDESMUS ACUTUS CALCULATED FROM

$$\log \frac{x + \alpha}{A - x} = \log \frac{\alpha}{A} + Kt; \text{ where } \alpha = 154, K = 0.071, A = 4550$$

| $t$ (DAYS) | $Kt$ | $Kt + \log \frac{\alpha}{A}$ | $m = \frac{x + \alpha}{A - x}$ | $Am$ | $Am - \alpha$ | $x$(CALC.) $= \frac{Am - \alpha}{1 + m}$ | $x$(OBS.) | $\theta$ |
|---|---|---|---|---|---|---|---|---|
| 3 | 0.213 | −1.2568 | 0.055 | 250 | 96 | 91 | | |
| 5 | 0.355 | −1.1148 | 0.077 | 350 | 196 | 182 | | |
| 6 | 0.426 | −1.0438 | 0.090 | 410 | 256 | 235 | 34 | −201 |
| 7 | 0.497 | −0.9728 | 0.107 | 487 | 333 | 300 | 148 | −152 |
| 8 | 0.568 | −0.9018 | 0.125 | 567 | 413 | 367 | 334 | − 33 |
| 10 | 0.710 | −0.7598 | 0.174 | 792 | 638 | 543 | 673 | 130 |
| 12 | 0.852 | −0.6178 | 0.241 | 1097 | 943 | 758 | | |
| 14 | 0.994 | −0.4758 | 0.334 | 1520 | 1366 | 1022 | 1091 | 69 |
| 19 | 1.349 | −0.1208 | 0.757 | 3444 | 3290 | 1873 | | |
| 23 | 1.633 | 0.1632 | 1.456 | 6625 | 6471 | 2640 | 2816 | 176 |
| 25 | 1.775 | 0.3052 | 2.019 | 9186 | 9032 | 3000 | 2992 | − 8 |
| 28 | 1.988 | 0.5182 | 3.298 | 15006 | 14852 | 3460 | 3348 | 112 |
| 31 | 2.201 | 0.7312 | 5.384 | 24497 | 24343 | 3815 | 3800 | − 15 |
| 34 | 2.414 | 0.9442 | 8.799 | 40035 | 39881 | 4080 | 4190 | 110 |
| 37 | 2.627 | 1.1572 | 14.36 | 65338 | 65184 | 4242 | 4350 | 108 |
| 40 | 2.840 | 1.4702 | 29.53 | 134362 | 134208 | 4400 | 4380 | − 20 |
| 42 | 2.982 | 1.5122 | 32.52 | 147966 | 147812 | 4400 | 4381 | − 19 |

The values of $x$ for corresponding values of $t$ were computed as follows:

Let $$\frac{x + \alpha}{A - x} = m$$

$$x + \alpha = Am - mx$$
$$x(1 + m) = Am - \alpha \qquad x = \frac{Am - \alpha}{1 + m}$$

The curve in figure 3 was drawn from the computed values of $x$. The root-mean-square deviation of calculated from observed values is 109. Since this is only 4.35 per cent of the mean value of $x$, the deviation is not extreme. The probable error of a single observation is 76.6.

The values of $K$, the growth constant, are greater at the beginning of a growth cycle than at any subsequent time if one computes $K$ from the observed values of $x$ and $t$ by means of equation (1). The decline in the values of $K$ seems to be a logarithmic function, not of time, but of the relative size of the organism.[5] Some of the advantages of equation (2) will be clear if we advert for the moment to one or two characters of equation (1). The fall in the value of $K$ results in a certain asymmetry of the growth

curve which is so pronounced in some cases that the form of the growth curve approaches that of a parabola, because the steepness of the curve depends upon the value of the constant, $K$. The initial fall in the values of the velocity constant may be regarded as proportional to the ratio $\frac{x+\alpha}{x}$.

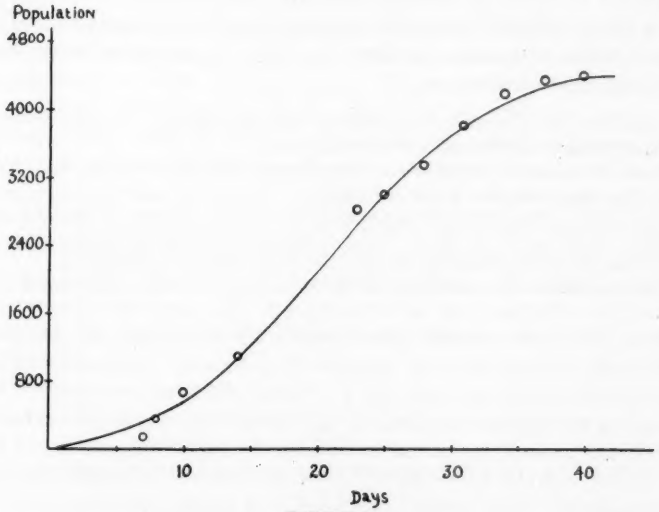



FIGURE 3

Growth curve of *S. acutus* calculated from the equation $\log \frac{x+154}{4550-x} = \log \frac{154}{4550} + 0.071t$. Small circles represent average number of individuals observed on the days indicated.

Thus we may write the differential forms of equations (2) and (1),

$$\frac{dx}{dt} = k'(x + \alpha)(A - x)$$

and

$$\frac{dx}{dt} = k(x)(A - x).$$

The values of $(A - x)$ in the two cases will differ so little that we may consider them equal, then

$$1 = \frac{k'(x+\alpha)}{kx}$$

and

$$\frac{k}{k'} = \frac{x+\alpha}{x}.$$

In the early life of the organism when $x$ is small the value of this fraction may be large. If $\alpha$ is small in comparison with $x$ the value of the ratio will rapidly approach unity and the decrease of $K$ will be observed only for the early values of $x$. However, if $\alpha$ is large with respect to $x$ the ratio will approach unity more slowly and the curve will have an increased asymmetry.

The use of equation (3) where a constant quantity, designated $\alpha$, is added to each value of $x$ seems justified, therefore, by quantitative as well as physiological considerations.

* Paper No. 264, University of California Graduate School of Tropical Agriculture and Citrus Experiment Station, Riverside, California.

[1] Reed, H. S., and R. H. Holland. "The Growth Rate of an Annual Plant, Helianthus," *Proc. Nat. Acad. Sci.*, **5**, 135–144 (1919).

[2] Reed, H. S., "The Nature of the Growth Rate," *J. Gen. Physiol.*, **2**, 545–561 (1920).

[3] Bray, W. C., and P. R. Davis. "The Autocatalytic Reduction of Bromate by Hydrogen Peroxide in Acid Solution," *J. Am. Chem. Soc.*, **52**, 1427–1435 (1930).

[4] I am grateful to Professor Bray for his interest and assistance contributed to this study.

[5] Reed, H. S., "Intra-seasonal Cycles of Growth," *Proc. Nat. Acad. Sci.*, **14**, 221–229 (1928).

---

## *STUDIES ON THE GROWTH HORMONE OF PLANTS. I*

By H. E. Dolk and K. V. Thimann

William G. Kerckhoff Laboratories of the Biological Sciences, California Institute of Technology

Communicated December 2, 1931

In recent years the process of cell elongation in plants has been studied by various investigators, Paal,[10] Went,[12] Cholodny,[3] and others. In these investigations it was proved that this process takes place under influence of a special substance, the so-called growth-substance ("Wuchsstoff"). Its physiological action, since it is assumed to be present in small quantities, is more or less comparable to that of the animal hormones. The presence of this substance in the coleoptiles of grasses was first established by Paal, and it was Went who extracted it from the tips of coleoptiles and who worked out a quantitative method for testing its activity. The hormone is not only found in higher plants, but similar substances are produced in relatively large quantities by various fungi, yeasts and bacteria (Nielsen[8,9], Boysen-Jensen[1,2]). The recent investigations of Heyn[7] have shown that the growth substance influences the cell elongation by changing the plasticity of the cell wall. In order to get a better idea

of the mechanism of its action, further knowledge about its chemical structure is necessary, and the following study is therefore a first attempt to establish the chemical nature of the substance.

§ 1. *Methods.*—The substance can be obtained from various sources. Its extraction out of the tips of coleoptiles did not appear very promising to us, as very large numbers of tips would have to be worked up in order to get a reasonable yield. At first an attempt was made to isolate it from malt syrup, as investigations of Seubert[11] have shown that this contains the substance in rather large amounts. However, it was soon found that the various commercial samples of malt give very different yields, some of them being quite inactive, probably through small differences in the method of preparation. It was therefore decided to use, as source of the substance, cultures of fungi. A pure culture of *Rhizopus suinus*, for which we have to thank Dr. Nielsen of Copenhagen, was used. In suitable media this mold produces large amounts of growth substance, as was shown by the careful and thorough studies of Nielsen. To what extent this substance is identical with that produced in the coleoptiles of grasses is uncertain, but this is not important for our purposes, as its physiological action is the same. According to Nielsen, the mold produces the hormone only when grown on solid media.* Under such conditions, however, the concentration of growth substance in the medium soon reaches a maximum and the production stops. The following arrangement was therefore made to avoid this.

The fungus was grown in large Petri dishes, 18 cm. in diameter, on filter paper soaked in nutrient medium. To one side a hole was drilled in the cover of the dish, through which by means of a syphon fresh sterile medium continuously dripped into the dish. On the other side, between cover and dish, was inserted a sterilized strip of filter paper, which acted as a syphon. By this arrangement a continuous stream of nutrient medium passed through the filter paper and in 24 hours about 150 cc. of active medium was collected from each dish. Under these conditions the fungus continues to produce the growth substance over a period of several days. In order to get still larger quantities of material this arrangement was later on altered in the following way. A large bell-jar with an opening in the top was filled with wood shavings soaked with nutrient medium. The opening was fitted with a stopper with two exit tubes, and the wide end of the bell-jar was covered with a glass plate in which four holes were drilled. The whole was inverted and sterilized in the autoclave, the shavings then inoculated with a suspension of the mold spores, and the apparatus placed in a room at 37° for two days to allow the mold to develop. After this time sterile nutrient medium was allowed to drip through the holes in the glass plate by means of syphons, and in order to keep the culture oxygenated, a current of sterile air was blown into the bell-jar

through one of the lower exit tubes. About 1 liter of medium dripped through in 24 hours. In all experiments, the medium described by Nielsen was used. Its composition is as follows: 10 g. dextrose, 10 g. ammonium tartrate, 0.5 g. $MgSO_4$, 0.5 g. $KH_2PO_4$, and a trace of $FeCl_3$, in 1 liter. In spite of all precautions, after some days the collected liquid often became infected with bacteria. In order to remove these the liquid was filtered through filter paper and a Berkefeld candle. The liquid was then concentrated under reduced pressure to about one-tenth of its volume, and stored at a temperature of 0° until required.

Its activity was tested on coleoptiles of *Avena sativa* by means of the method described by Went. In this method the active liquid is mixed with agar, small blocks of the agar are put on one side of the decapitated plants, and the difference in growth between the two halves of the coleoptiles measured by means of the curvature produced. A pure line of "Sieges hafer," obtained through the kindness of Dr. Åkermann of Svalöf was used.

The seeds, deprived of their husks, were soaked in water for four hours, placed on filter paper for 24 hours to germinate, and planted in glass holders as described by Went. The seedlings were grown for 48 hours in a dark room at a constant temperature of 25° and a relative humidity of 85%. The coleoptiles were then decapitated at 5 mm. from the tip, the primary leaf pulled loose, and after 40 minutes the agar blocks were put on one side of the stump by means of a drop of 10% gelatin. 110 minutes after the agar was put on, the curvatures were photographed by making a shadow print on bromide paper. To wait any longer than 150 minutes between decapitating and photographing is not advisable, since the formation of the new physiological tip would interfere with the test (Dolk[4]).

The agar blocks were prepared by mixing 1 cc. of active liquid with an equal volume of 3% agar, and pouring $^1/_2$ cc. of the mixture into a brass ring 20.6 mm. in diameter. After setting, a small rectangular plate 8.0 mm. × 10.7 mm. was cut out and this divided into 12 equal smaller blocks, special tools similar to those already described by Dolk[5] being used. Each resulting block had a volume of 10.7 $mm.^3$ As was found by Went, and confirmed by Nielsen, up to a certain concentration the curvature is directly proportional to the amount of active material present in the agar. Above this concentration it becomes less and less, and reaches finally a constant value, ("maximum angle"). In our case it was found that up to an angle of 14° there was direct proportionality, therefore in testing the active material it was always diluted beforehand to such a concentration that the angle obtained was between 5° and 14°. Each test was made on 8–12 plants to get a satisfactory mean value.

The curvatures obtained were determined by measuring the angle between the uppermost part of the stump and the uncurved base, with a

tangimeter. This gives a direct measure of the difference in growth between the two sides of the coleoptile, providing the thickness does not vary considerably. As was shown by Went, the thickness of this type of oat varies only between very narrow limits. This method has the advantage that only one measurement has to be made, while that used by Nielsen and by Boysen-Jensen involves at least two, i.e., the length of the curved part and the radius of curvature. Small growth differences, especially, cannot be determined accurately by this method, since radii longer than 8 cm. have to be neglected. This is presumably the reason why the curves relating reaction of the plant to concentration of growth substance, given by Nielsen, do not pass through the origin, but meet the abscissa at a point some distance away, which should mean that no reaction is produced by small concentration of growth substance. This is, of course, not the case, since Went and also van der Wey[13] found for the same relation a straight line passing through the zero point.

In order to be able to bring all the data on to a quantative basis, some kind of a unit system had to be introduced. As *unit* was chosen that quantity of growth substance which has to be present in 1 cc. of solution to give, after mixing with 1 cc. of agar, an angle of 1°. The total number of units per cc. of solution is then found by multiplying the angle measured by the dilution in which the test was carried out. In the experiments below, the first figure gives always the dilution, the second the angle actually measured. The actual amount of material in the block applied to the plant is only $^1/_{200}$ of that which is present in 1 cc., and this quantity is termed a "plant unit." The units defined above have only an arbitrary value, since their application is limited to measurements in which the procedure described above is rigidly adhered to. The same dependence upon the conditions of the test applies to the biological assay of all animal hormones. The variety of oat used, the culture conditions, the size of the agar blocks and the time relations between the various manipulations, will all exert a considerable influence upon the final result. Van der Wey, for instance, showed that variations in the time relations alone can alter the curvatures by as much as 10 times. This is the reason that the unit as defined above cannot be compared with that proposed by Boysen-Jensen. Duplicate determinations on the same solution, however, showed that if the standardized conditions are followed, the closest possible agreement can be obtained.

In order to obtain active solutions of growth substance, the concentrated culture media were always extracted with ether to get rid of most of the salts and impurities, since Nielsen found the growth substance to be soluble in ether. The most efficient extraction can be obtained by equilibrating the solution three times with half its volume of ether. Most of the ether was distilled off, water was added, and the solution boiled to remove the

last traces of ether. As already pointed out by Nielsen, great care has to be taken to use pure ether, since traces of peroxides readily destroy the growth substance. At the beginning, ether freshly distilled over sodium was used, but even with this, considerable destruction of growth substance took place, probably due to traces of peroxides still remaining. Various methods of purifying the ether were therefore tried out:

(1) One hundred and fifty cc. ether (Merck blue label) was kept overnight over fresh sodium wire, and then distilled. Two cc. of solution, with an activity of $16 \times 12.0° \times 2$ cc. = 384 units, + 3 cc. water, were shaken well with 5 cc. of this ether, the ether layers mixed with 10 cc. water and the ether boiled off. Activity of the ether extract, $3 \times 9.0° \times 10$ cc. = 270 units; activity of the water layer; $5.5° \times 5$ cc. = 28 units. Total activity, 298 units, hence loss = 20.4%. In most of the extractions made with ether so purified, a similar loss of about 20% was found.

(2) One hundred and fifty cc. ether was shaken well with an equal volume of 30% NaOH solution, and washed five times with distilled water to remove the NaOH. Two cc. of the same solution as above + 3 cc. water were shaken in the same way with 5 cc. ether. Activity of ether layer, $3 \times 9.1° \times 10$ cc. = 273 units; activity of the water layer, $6.5° \times 5$ cc. = 33 units. Total activity; 306 units, hence loss = 20.0%.

(3) One hundred and fifty cc. ether were allowed to stand overnight with 7 g. $FeSO_4 \cdot 7H_2O$ + 1 g. CaO and then distilled. Two cc. of the same solution + 3 cc. water were shaken with 5 cc. ether as above. Activity of ether layer, $3 \times 10.2° \times 10$ cc. = 306 units. Activity of water layer, $6.5° \times 5$ cc. = 33 units. Total activity, 339 units, hence loss = 11.5%.

From these experiments it is clear that if ether is purified with ferrous sulphate and lime (Garbarini's method[6]), the best results are obtained. A small loss was still found with this ether, but this is at any rate partly due to mechanical loss on the walls of the vessel.

For the determination of the total dry matter in the solution, $^1/_2$ cc. was pipetted by means of an Ostwald pipette into a small platinum dish. This was placed on top of a small Erlenmeyer flask containing boiling redistilled water, and evaporated to dryness. It was then transferred to the oven at 110° for five minutes, cooled in a micro-desiccator and weighed on the micro-balance. Since the amounts available for weighing were very small, the error was somewhat high, but duplicate determinations usually agreed within 2–3%, when the total amount weighed was 0.2 mg. or more. After weighing, the dish and contents were heated to redness for a moment, cooled in the desiccator and again weighed to obtain the ash. For the loan of a Christian Becker micro-balance the authors are greatly indebted to Dr. Chester Bliss of the U. S. Bureau of Entomology at Whittier, Calif.

§ 2. *Determination of the Partition Coefficient and Dissociation Constant.*— In the preliminary experiments on the effect of pH, it was found that the active solutions of the growth substance were somewhat acid, having a pH of about 4. On the addition of alkali and extraction of the alkaline solution with ether, the ether extract was found to be inactive. The growth substance was not destroyed by the alkali, for the remaining aqueous layer was still highly active.

To 25 cc. active solution was added 0.1 M $Ba(OH)_2$ to a pH about 10 (colorimetric), 1.0 cc. being required. The solution was extracted three times with 12–13 cc. freshly redistilled ether. Activity of the original solution $8 \times 6.8° \times 25$ cc. = 1360 units. Ether extract $8 \times 0° = 0$ units. Aqueous layer $16° \times 25$ cc. = > 400 units. (max. angle). A second experiment gave identical results.

It is clear from the above that formation of a salt has occurred with an accompanying loss of solubility in ether. The growth substance must therefore be considered as an acid. The pH limits within which salt formation occurs were determined as follows: to 5 cc. of an active solution were added 5 cc. McIlvaine phosphate-citrate buffer of pH 4, 5, 6 and 7. The mixture was extracted 3 times with 5 cc. freshly distilled ether and the extract made up to 10 cc. Activities of extracts:

| | |
|---|---|
| pH 4 | $2 \times 9.6° \times 10$ cc. = 192 units |
| pH 5 | $2 \times 9.7° \times 10$ cc. = 194 units |
| pH 6 | $2 \times 5.7° \times 10$ cc. = 114 units |
| pH 7 | $1 \times 0° \times 10$ cc. = 0 units |

It should be added that confirmatory tests on the remaining aqueous layers were carried out, but for reasons which are given below, did not give satisfactory curvatures, owing to the effect of the buffer on the test. The region of pH concerned is from 5 to 7. It is clear that by narrowing the range, the actual dissociation constant of the acid could be found, and this was the next step.

In order to determine accurately the dissociation constant of the growth substance, however, account must be taken of the fact that the extraction with ether displaces the equilibrium between the free acid and its salt. The amount by which the equilibrium is displaced depends on the amount of the acid which is extracted, and therefore on its partition coefficient between ether and water. This constant must therefore be determined in order to calculate the dissociation constant of the acid.

If the growth substance is an acid, then

$$\frac{[H][A]}{[HA]} = K \qquad (1)$$

where $K$ is the dissociation constant. If $HA$, the free acid, is the ether-soluble form, and $T$ the total amount of growth substance, then we have

$$\frac{[HA]}{[A]} = \frac{[H]}{K}$$

and hence

$$\frac{[HA]}{[A] + [HA]} = \frac{[H]}{[H] + K}. \tag{2}$$

Since $[A] + [HA] = [T]$, the left-hand side of (2) is $[HA]/[T]$, the fraction of growth substance extractable with ether, which can be written $F$, i.e.,

$$F = \frac{[H]}{[H] + K} \tag{3}$$

or

$$K = \frac{[H]\,(1 - F)}{F} \tag{3a}$$

The partition coefficient, $P$, is defined as the ratio of $HA$ in ether and in water at equilibrium, i.e.,

$$P = \frac{[HA]_e}{[HA]_w}. \tag{4}$$

If the volumes of ether and water are the same, concentrations can be replaced by amounts. After one extraction with an equal volume of ether,

then

$$T = HA_w + A_w + HA_e. \tag{5}$$

From (1)

$$[A] = \frac{K[HA]}{[H]}.$$

Replacing concentrations by amounts, and considering only the water phase,

$$A_w = \frac{K\,HA_w}{H_w}.$$

Equation (5) therefore becomes

$$T = HA_w\left(1 + \frac{K}{H_w}\right) + HA_e$$

or from (3)

$$T = HA_w\left(\frac{1}{F}\right) + HA_e$$

Substituting from (4)

$$T = \frac{HA_e}{PF} + HA_e$$

whence $$\frac{HA_e}{T} = \frac{PF}{1 + PF} \tag{6}$$

or $$F = \frac{HA_e}{P(T - HA_e)}. \tag{6a}$$

The left-hand side of (6) is the fraction of the total growth substance to be found in the ether extract after one equilibration with an equal volume of ether. If half the volume of ether is used,

$$\frac{HA_e}{T} = \frac{PF}{2 + PF}$$

By similar treatment, with an initial concentration of $[HA]_w$, the equation for the second extraction was developed, and also for the third. These, in the case of half the volume of ether, were found to be, respectively,

$$\frac{HA_{e_2}}{T} = \frac{2PF}{(2 + PF)^2}$$

and $$\frac{HA_{e_3}}{T} = \frac{4PF}{(2 + PF)^3}.$$

The total fraction extracted by three equilibrations with half the volume of ether is obtained by adding

$$\frac{HA_{e_1} + HA_{e_2} + HA_{e_3}}{T} = \frac{PF}{2 + PF}\left(1 + \frac{2}{2 + PF} + \frac{4}{(2 + PF)^2}\right). \tag{7}$$

However, no greater accuracy is obtained by several extractions, and a single extraction was therefore used in all those experiments from which $P$ and $pK$ were calculated.

*Partition Coefficient.*—For the determination of the partition coefficient, the following experiments were carried out. To 2 cc. active solution were added 3 cc. water, and the solution shaken thoroughly with freshly distilled ether. The experiments were carried out in the cold room at 1° to minimize evaporation of the ether. Water was added to the ether layer, and both ether and aqueous layers were then boiled until free from ether, made up to volume, and the activities of both determined. The results are shown in table 1.

*Dissociation Constant.*—For these experiments a measured volume of an active solution was pipetted into a slightly larger volume of Sørensen's M/15 phosphate buffer, the total volume in each case being 5 cc. The activity of the original solution having first been determined, the number

of units per cc. of the mixture of buffer and growth substance was thence obtained. (Column 5 of table 2.) The mixture was then equilibrated with 5 cc. of ether freshly distilled from $FeSO_4$ + CaO and sodium wire, the operation being carried out in the cold room at 1° as before. Finally 3 cc. water was added to the ether extract, the ether boiled off, the solution made up to 5 cc. and the activity determined. On account of the modifying effect of the buffer on the activity tests (see below), only the ether extract could be tested. In the more acid solutions, where the extraction was more or less complete, the value of $T - HA_e$ in equation (6a) becomes

TABLE 1

PARTITION COEFFICIENT OF GROWTH SUBSTANCE BETWEEN ETHER AND WATER AT 1°C.

| ACTIVITY OF ETHER EXTRACT | FINAL VOL., CC. | TOTAL ACTIVITY IN ETHER, UNITS | ACTIVITY OF AQUEOUS RESIDUE | FINAL VOL., CC. | TOTAL ACTIVITY IN WATER, UNITS | $P$ |
|---|---|---|---|---|---|---|
| 12.0° | 20 | 240 | 2.6° | 10 | 26 | 9.2 |
| 12.8° + 12.9° | 10 | 128 + 129 = 257 | 2.7° | 10 | 27 | 9.5 |
| 3 × 10.2° | 10 | 306 | 6.5° | 5 | 33 | 9.3 |
| 3 × 9.0° | 10 | 270 | 5.5° | 5 | 28 | 9.6 |

Mean value of $P$ = 9.4.

a small difference between two large numbers. Since this is then divided into a large number, $HA_e$, the error becomes large. On this account most of the measurements were made toward the alkaline side of the dissociation curve. One point, however, was determined at rather high acidity, and this was in general agreement with the others. The buffer for this point was an improvised $KH_2PO_4$-HCl mixture, whose pH was determined with the hydrogen electrode. The value obtained from this experiment was not used in determining the mean, for the reasons given

TABLE 2

DISSOCIATION CONSTANT OF GROWTH SUBSTANCE

| I pH OF BUFFER | II VOLUME OF BUFFER SOLUTION, CC. | III ACTIVITY OF ORIGINAL SOLUTION, UNITS/CC. | IV VOLUME TAKEN, CC. | V ACTIVITY OF MIXTURE = $[T]$, UNITS/CC. | VI ACTIVITY OF ETHER EXTRACT = $[HA_e]$ UNITS/CC. | VII FRACTION OF G. S. IN $HA$ FORM = $F$ | VIII DISSOC. CONSTANT OF G. S. = $K$ ($\times 10^5$) | IX $-$LOG, DISSOC'N. CONSTANT = $pK$ |
|---|---|---|---|---|---|---|---|---|
| 4.51 | 4 | 16 × 12° = 192 | 1 | 38.4 | 4 × 8.4° = 33.6 | 11.1 % | 1.07 | 4.97 |
| 5.29 | 3 | 96 × 5.7° = 547 | 2 | 219 | 16 × 9.0° = 144 | 74.5 % | 1.99 | 4.70 |
| 5.59 | 4 | 192 | 1 | 38.4 | 2 × 9.8° = 19.6 | 20.4 % | 1.99 | 4.70 |
| 5.91 | 4 | 192 | 1 | 38.4 | 1.5 × 10.2° = 15.3 | 7.05% | 1.82 | 4.74 |
| 6.25 | 4 | 192 | 1 | 38.4 | 1 × 8.9° = 8.9 | 3.22% | 1.70 | 4.77 |
| 6.47 | 4.5 | 547 | 0.5 | 54.7 | 1.5 × 6.0° = 9.0 | 2.10% | 1.58 | 4.80 |
| 6.64 | 4.5 | 547 | 0.5 | 54.7 | 1 × 6.0° = 6.0 | 1.32% | 1.74 | 4.76 |
| 6.98 | 3 | 547 | 2 | 219 | 1 × 10.5° = 10.5 | 0.53% | 1.95 | 4.71 |
| | | | | Mean of all except first determination | | | 1.82 | 4.75 |

above. Since the volumes of the initial mixture and of the ether were the same (5 cc.), the ratio of column 6 over column 5 gives at once the value of $HA_e/T$. The values of $F$ in column 7 were obtained from equation (6a), using 9.4 as the value of $P$. The calculation of $K$ from equation (3a), and hence of $pK$, follow in columns 8 and 9. It should be noted that the $pK$ is for solutions containing up to $M/15$ phosphate.

The fraction dissociated at varying pH is plotted in figure 1, in which the curve is the theoretical dissociation curve for an acid of $pK$ 4.75.

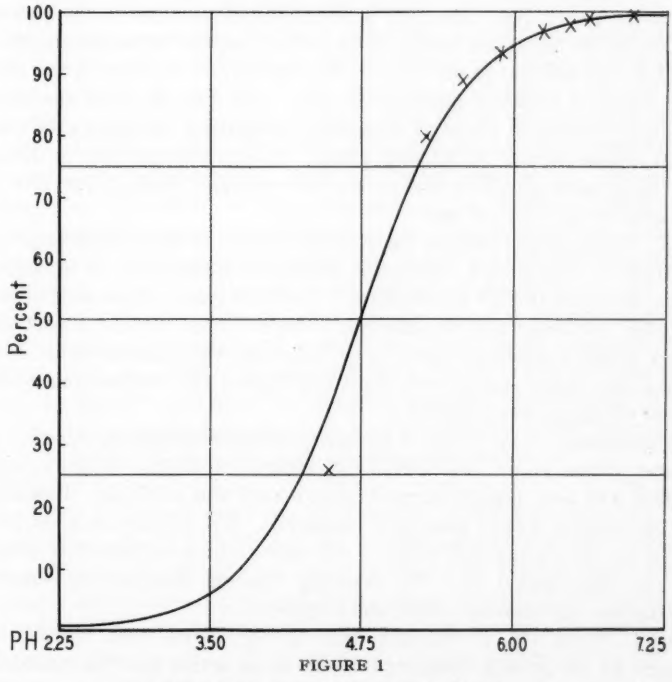



FIGURE 1

Dissociation curve of the growth substance.

The fact that the growth substance is only extractable with ether in acid solutions is very important. When media in which various organisms have grown are extracted in order to test their activity, great care has to be taken to adjust the pH beforehand, or else serious errors can be introduced. For the same reason, it was important to determine whether the growth substance in salt form had the same effect on the plant as in the free acid form. In the first experiments to determine the $pK$, the activity both of the water and of the ether layer were tested, as men-

tioned above. At pH 6 and 7 the activity of the water layer was in each case too low, which indicated that the state in which the growth substance was present affects the tests. In order to establish this further, the following experiments were carried out. 0.5 cc. growth substance solution of an activity 32 × 12.6° × 0.5 cc. = 202 units, were diluted 1:32 and 0.5 cc. of this solution was mixed with 0.5 cc. water and 1 cc. agar. Activity found: 32 × 12.6° × 0.5 cc. = 202 units. Another 0.5 cc. mixed with 0.5 cc. $M/15$ phosphate buffer pH 7.0 and 1 cc. agar gave activity 32 × 3.3° × 0.5 cc. = 53 units. Hence the presence of the buffer seriously affects the test. That this effect is due to the pH and not to the high salt content of the buffer can be shown by adding, instead of phosphate, $M/15$ NaCl to the solution to be tested. The activity in this and in the blank were identical within experimental error. The pH effect on the test is principally found in buffered solutions; unbuffered solutions give much better results even in the alkaline range. This is probably due to the fact that the growth substance will be rapidly converted from its salt into the free acid by the acid cell-sap.

§ 3. *Chemical Properties of the Growth Substance.*—The outstanding characteristic of the growth substance, beside its acidity, is, as mentioned above, its susceptibility to oxidation. Since the traces of peroxide present in freshly distilled ether are sufficient to destroy some 20% of the activity, strong positive reactions were to be expected with laboratory oxidizing agents, and experiments were therefore carried out to determine what type of oxidation was involved.

*Permanganate.*—To $^1/_2$ cc. of an active solution containing 79 units per cc. was added $^1/_2$ cc. of a 0.0024 $M$ solution of alkaline $KMnO_4$. The solution was very gently warmed, then cooled and acidified. Reduction had taken place, a faint pink color remaining. The solution was extracted with 4 cc. freshly distilled ether, 1 cc. water added to the ether extract and the ether boiled off. The resulting solution was entirely inactive. Destruction was therefore rapid and complete.

*Silver Nitrate.*—A preliminary test having shown that $AgNO_3$ was reduced by the growth substance, $^1/_2$ cc. of an active solution containing 180 units per cc. was treated with $^1/_2$ cc. of 0.5 $M$ $AgNO_3$, i.e., a considerable excess, and warmed in hot water until precipitation of Ag was complete. The solution was cooled, acidified and extracted with 5 cc. freshly distilled ether, the extract freed from ether and tested as before. A very slight angle, 10 × ?1° = ?10 units per cc., indicated that the activity was almost if not quite destroyed. The oxidation is probably somewhat slower with $AgNO_3$ than with $KMnO_4$.

*Fehling's Solution.*—Since the silver reaction is principally given by aldehydes, oxidation by Fehling's solution was tried. In this case a blank to compensate for the possible effect of the alkali was necessary.

To $^1/_2$ cc. of a solution containing 59 units per cc. was added $^1/_2$ cc. each of the two solutions comprising the Fehling's mixture, and as a blank, to another $^1/_2$ cc. was added $^1/_2$ cc. of the alkaline tartrate solution alone and $^1/_2$ cc. water. Both tubes were placed in boiling water for 10 minutes, cooled, acidified with HCl to pH 4–5, and extracted once with 5 cc. freshly distilled ether. The ether extracts were freed from ether and tested. Activity of solution treated with copper and alkali: $2 \times 17.4° = > 35$ units per cc. (max. angle); activity of solution treated with alkali only; $2 \times 19.6° = > 39$ units per cc. (max. angle). Hence the growth substance is not oxidized by Fehling's solution.

*Oxime.*—An experiment to test whether any destruction was caused by hydroxylamine gave inconclusive results. To $^1/_2$ cc. of a solution containing 120 units per cc. was added 1 cc. of a strong solution of $NH_2OH.HCl$ and after standing with gentle shaking the solution was extracted with 3 cc. freshly distilled ether and both extract and aqueous layer were tested. Activity of the ether extract: $12 \times 7.8° = 94$ units per cc. Activity of water layer: $8 \times 0.7° = 5.6$ units per cc. It is, however, possible that the equilibrium between the substance and its oxime, if formed, is displaced by the ether extraction in the same way as the equilibrium between the acid and its salt. Or, since most oximes are soluble in ether, the plant may be able to reconvert the oxime to the original growth substance. It may, however, be concluded that no definite evidence was obtained to indicate that the growth substance is an aldehyde. The extreme sensitivity to $KMnO_4$ and $AgNO_3$ indicates, however, that double linkings of some sort must be present. It is hoped to obtain further light on the nature of the double bond or bonds in subsequent work.

*Effect of Acid and Alkali.*—The stability with regard to acid was tested in the cold and in the hot. To $^1/_2$ cc. of an extract containing 180 units per cc. was added $^1/_2$ cc. of 2 *N* HCl. After standing about 3 hours at room temperature the solution was extracted once with an equal volume of ether and the activity of the ether extract tested. This should take out 90% of the active substance. In case the high acidity should have affected the test, another $^1/_2$ cc. similarly treated with HCl was first neutralized with NaOH and then an equal volume of pH 6 buffer added. Activity of the first extract $20 \times 5.1° = 102$ units per cc. Activity of the 2nd extract $20 \times 4.6° = 92$ units per cc. In each case the activity is reduced by about a half through the acid treatment.

Another $^1/_2$ cc. of the same extract was heated in boiling water 30 minutes with $^1/_2$ cc. 2 *N* HCl, cooled and extracted once with an equal volume of ether. From the above experiment it was clear that the 1 *N* HCl was not extracted sufficiently in the ether to affect the test, hence the acid solution was extracted directly. Activity of original solution $20 \times 8.7° = 174$ units per cc. Activity of acid extract $10 \times 1.1° = 11$ units per cc. Hence

destruction by 1 *N* acid proceeds rather slowly, requiring more than 30 minutes at 100° and more than three hours at 25°.

It was mentioned above that a blank was carried out for the alkali of the Fehling's solution, indicating that no effect occurred. However, a further experiment was carried out to confirm this. To one cc. of an active extract was added 1 cc. 2 *N* NaOH, and the liquid heated 20 minutes in boiling water, cooled, acidified with $H_2SO_4$ and extracted with three volumes freshly distilled ether. The ether extract was boiled and the activity tested as before. A similar tube containing 1 cc. of the same active extract and 1 cc. water, heated for the same time, acidified and extracted, served as blank. Activity of alkali heated extract; $5 \times 18.5° = 92$ units per cc. Activity of blank $5 \times 17.8° = 89$ units per cc. Hence no destruction is caused by 1 *N* NaOH in 20 minutes at 100°.

§4. *Experiments on Purification. Dialysis.*—Various attempts were made to get the substance free from impurities. This was first tried by means of dialysis. As Nielsen had shown that on dialysis through a collodion membrane the substance is destroyed, presumably by peroxides still present, a cellophane membrane was used.

TABLE 3

RESULTS OF DIALYSIS

| | AFTER 24 HOURS | AFTER 3 × 24 HRS. | | |
|---|---|---|---|---|
| SOLUTION | ACTIVITY PER CC. | ACTIVITY PER CC. | DRY WEIGHT PER CC. | DRY WEIGHT PER UNIT |
| Inner solution | 42 units | 8.8 units | 0.45 mg. | 0.051 mg. |
| Outer solution | 6 units | 8.0 units | 0.49 mg. | 0.061 mg. |

A piece of glass tubing, 15 mm. bore was ground at one end and fitted into a rubber stopper so that the ground end was just flush with the surface of the stopper. Over the tube and stopper a cellophane membrane was stretched and fastened with rubber bands. Ten cc. of growth substance solution was brought into the tube and dialyzed against 25 cc. redistilled water. The inner and outer solutions were stirred by bubbling nitrogen through, and dialysis allowed to take place in the cold room at 1°. After 3 days, equilibrium was reached and the solutions were tested. A test with carmine solution, after the experiment, showed the membrane to be intact.

A second experiment gave similar results.

Hence the activity as well as the total dry matter per cc. is about equally distributed over both solutions, so that in this way no appreciable purification can be obtained.

*Evaporation.*—After extracting the active medium with ether, evaporating off the ether and taking up in water, Nielsen observed that the growth

substance was more soluble than the rest of the oil and obtained a partial separation by this means. In our experiments the whole of the oil was dissolved in water, but on cooling the solution a large part of it again separated. Thus, in one extraction, the aqueous solution contained $96 \times 5.7° = 547$ units per cc. or 13,700 units in the whole volume of 25 cc. The oil which separated on cooling was redissolved in chloroform and taken up in 10 cc. water, when part of it was dissolved, to give $8 \times 7.4° = 59$ units per cc. or 590 units in the whole 10 cc. The remaining oil was again dissolved in chloroform and taken up in 10 cc. water to give a solution containing $1 \times 3.9° = 3.9$ units per cc. or 39 units in 10 cc. Hence the whole of the oil, representing the bulk of the solid matter, contained only 631 units, while the aqueous solution contained 13,700 units. On this basis it was hoped to achieve further purification of the aqueous solution by evaporation and cooling. For this purpose 16 cc. of active solution containing 293 units per cc. was evaporated to about 1 cc., cooled to 1°C.

TABLE 4

| | I SEPARATED OIL DISSOLVED IN CHLOROFORM | | II AQUEOUS REMAINDER | III AQUEOUS REMAINDER REEXTRACTED WITH ETHER |
|---|---|---|---|---|
| Activity (units/cc.) | $24 \times 10.4° = 250$ | | $24 \times 5.0° = 120$ | $8 \times 9.7° = 77.6$ |
| Dry weight (mg./cc.) | 0.78 | 0.82 | 0.59 | 0.123 |
| Ash (mg./cc.) | 0 | 0.04 | 0.21 | 0.001 |
| Dry weight organic matter only | 0.78 | 0.78 | 0.38 | 0.122 |
| Weight 1 unit/cc. (mg.) | 0.00312 | | 0.00317 | 0.00157 |
| Weight 1 unit on plant (mg.) | $15.6 \cdot 10^{-6}$ | | $15.8 \cdot 10^{-6}$ | $7.8 \cdot 10^{-6}$ |

and the separated oil skimmed off the surface with a fine glass rod and redissolved in chloroform. Water was added, the $CHCl_3$ evaporated off and the solution made up to 10 cc. The 1 cc. of clear yellow aqueous solution was made up to 10 cc. again. The activities and dry matter of the two solutions were then determined, I and II in table 4. Since the aqueous solution carried all the dissolved salts present, the ash was 36%. The weight of organic matter was corrected for this ash, since the fact that active solutions could be obtained without ash indicates that the growth substance is wholly organic. However, this correction is necessarily a minimum one, since the metallic impurity which is determined as ash probably has a higher weight when present in undecomposed form.

A comparison of the dry weight per unit indicates that these two solutions were of identical purity and therefore no purification resulted. The separation of oil which took place in the early stages of the preparation must therefore have removed all of the less water soluble oil. It may be deduced from this experiment that the solubilities in water of the

growth substance and its inactive contaminants at the later stages are the same.

Lastly the remainder of the aqueous solution was extracted three times with an equal volume of ether and made up back to the same volume with water. Though the activity of this solution had fallen to about $^2/_3$, the dry weight had fallen still more and a considerable purification resulted. (III in table 4.) Though it is not possible, therefore to purify the extract further by simple evaporation and chilling, yet reëxtraction with ether is effective, probably on account of the high partition coefficient.

*Fractional Extraction.*—It follows from the value of the dissociation constant that at pH 7 only a very small fraction of the growth substance will be extracted with ether. It can be calculated from the formula, equation (7), that the total fraction extracted by three times equilibrating with $^1/_2$ the volume of ether will be 6.7%. Any other substance present, unless also an acid of similar strength, will however be removed in the ether, and the ratio of activity to dry matter in such an extract should be lower if impurities are present. The usefulness of this method for purification was therefore examined as follows: two cc. of an active

TABLE 5

| SOLUTION | ACTIVITY UNITS/CC. | TOTAL UNITS | DRY WEIGHT MG./CC. | DRY WEIGHT MG./UNIT |
|---|---|---|---|---|
| pH 6.98 ext | $2 \times 5.3° = 10.6$ | 65.7 = 11% | 0.159 | 0.0150 |
| pH 5.29 ext. | $16 \times 6.4° = 102.4$ | 512 = 88% | 0.426 | 0.00416 |

solution containing $24 \times 12.2° = 293$ units per cc. were mixed with 4 cc. Sørensen's phosphate buffer of pH 6.98. The McIlvaine buffer could not be used owing to the solubility of citric acid in ether. The mixture was thoroughly shaken three times with 3 cc. of freshly distilled ether in the cold at 1°C., the extract freed from ether, made up to 6.2 cc. and the activity and dry weight determined. An exactly similar experiment was then carried out using buffer of pH 5.29, at which pH about 86% should be extracted. The results, given in table 5, show that there is a considerable difference in the activity : dry weight ratio.

By subtracting the activities and dry weights of the pH 7 extract from those of the pH 5.3 extract it is clear that the 91.8 units per cc. difference corresponds to a weight of 0.267 mg. per cc., or 1 unit is 0.00291 mg. Hence if everything is assumed to be extracted at pH 5.3, then by first extracting at pH 7 and then acidifying and reëxtracting, an increase in the purity given by $(1/0.00291 - 1/0.00416)0.00416 = 42.7\%$ would be obtained.

This method was therefore applied to a considerable amount of highly active solution. This solution was prepared in the ordinary way by ether extraction of active culture medium, dissolving in water, thoroughly

chilling, and then decanting from separated oil. It contained $64 \times 4.5° = 288$ units per cc. and was already of high purity, the dry weight being 0.52 mg. per cc., so that 1 unit per cc. = 0.0018 mg. per cc., or 1 plant unit = $9.10^{-6}$ mg. 21.5 cc. of this solution were mixed with pH 6.98 phosphate buffer and extracted three times with 25 cc. freshly distilled ether at 1° C. The extract was freed from ether and made up to 10 cc. The aqueous solution was acidified with HCl to a pH about four (colorimetric) and again extracted three times with an equal volume of ether. The activity and dry weight of both extracts were determined. A portion of the pH 4 extract was then reëxtracted three times with $^1/_2$ the volume of ether and the activity and dry weight again determined. This agreed satisfactorily, showing approximately a constant weight per unit. Where duplicate determinations were carried out, both are given in the table.

TABLE 6

| | ACTIVITY | | DRY WEIGHT | | | | DRY WEIGHT OF 1 UNIT CALCULATED ON ASH-FREE BASIS | |
|---|---|---|---|---|---|---|---|---|
| SOLUTION | UNITS/CC. | TOTAL UNITS | MG./CC. | ASH | CORRECTED MG. | TOTAL MG. | MG./UNIT | MG./PLANT UNIT |
| Original Solution *A* | $64 \times 4.5° = 288$ | 6180 | 0.52 | 2% | 0.51 | 11.2 | 0.0018 | $9{\cdot}10^{-6}$ |
| Solution *A*, extracted at pH 7 = *B* | $8 \times 9.9° = 79.2$ | 792 | 0.388 | 5% | 0.370 | 3.9 | 0.0049 | $24.5{\cdot}10^{-6}$ |
| *A* then extracted at pH 4 = *C* | $32 \times 13.2° = 423$<br>$64 \times 6.3° = 403$ | 4130 | 0.712 | 0 | 0.712 | 7.1 | 0.00172 | $8.6{\cdot}10^{-6}$ |
| Part of sol. *C* reëxtracted with ether = *D* | $8 \times 10.0° = 80$<br>$16 \times 5.0° = 80$ | 800 | 0.155<br>0.156 | 8% | 0.142 | .. | 0.00178 | $8.9{\cdot}10^{-6}$ |
| Part of *D* reëxtracted with chloroform | $6 \times 10.3° = 61.8$ | 309 | 0.089<br>0.084 | 0 | 0.0865 | .. | 0.00139 | $7.0{\cdot}10^{-6}$ |

Comparison of the total dry weights of solutions B and C with that of the initial solution A shows that all the solid material is accounted for within 2%, so that no loss has occurred. The total activities show, however, that inactivation to the extent of about 20% has occurred. Unless this inactivated material is quantitatively extracted at pH 7, which is not likely, some of it will be present in the pH 4 extract. Therefore, although some impurity has certainly been removed in the pH 7 extract, the net increase in purity is only slight. Some inactivation appears to be inseparable from each ether extraction, so that the highest purity could never be expected to be reached by reëxtracting with ether. Accordingly a final extraction of a part of the last extract was made with chloroform, using two equilibrations with an equal volume of chloroform. By this method the final dry weight per plant unit was found to be 7.0 millionths of a milligram, which represents only a slight improvement over the purity of the final ether extractions above, and also of that in table 4. It seems probable from the above data that we have here to deal with a

mixture of the growth substance and its oxidized form, the two substances being probably so closely related that even fractional extraction at different H-ion concentrations does not separate them.

For further progress in purification, larger quantities of material will be necessary in order to obtain sufficient for the preparation of crystalline derivatives and for analysis. Experiments along these lines are already in progress.

## SUMMARY

1. A technique is described by which it is possible to obtain, from cultures of molds, considerable quantities of the substance which controls the elongation of plant cells.

2. The activity of this "growth substance" is tested by its action upon coleoptiles of Avena, and a biological system of units is suggested, whereby, using a standardized technique, it is possible to express the activities of the substance quantitatively.

3. The growth substance is proved to be an acid, which has a dissociation constant of $1.8 \cdot 10^{-5}$.

4. Some of the chemical properties, in regard to its oxidizability and stability to acids and alkalies have been determined.

5. A method for purification is given, by which extracts of approximately constant purity can be obtained.

6. The highest purity obtained corresponded to $7.0 \cdot 10^{-6}$ mg. dry material per plant unit.

* Unpublished work by Mr. Bonner, subsequently carried out in this laboratory, has shown that under suitable conditions *Rhizopus* does produce large amounts of the hormone in liquid media.

[1] Boysen-Jensen, *P. Bioch. Zeitschr.*, **205**, 236, 1931.
[2] Boysen-Jensen, *Bioch. Zeitschr.*, **1**, 238, 1931.
[3] Cholodny, N., *Biol. Zentralbl.*, **47**, 604, 1927.
[4] Dolk, H. E., *Proc. Kon. Akad. Wetensch., Amsterdam*, **29**, 1113, 1926.
[5] Dolk, H. E., *Geotropie en Groeistof*, Diss. Utrecht, 1930.
[6] Garbarini, *Chem. Zentralbl.*, 1126, 1909.
[7] Heyn, A. N. J., *Rec. Trav. bot. neérl.*, **28**, 113, 1930.
[8] Nielsen, N., *Jahrb. wiss. Bot.*, **73**, 125, 1930.
[9] Nielsen, N., *Bioch. Zeitschr.*, **23**, 244, 1931.
[10] Paal, A., *Jahrb. wiss. Bot.*, **58**, 406, 1918.
[11] Seubert, E., *Z. f. Bot.*, **17**, 49, 1925.
[12] Went, F. W., *Rec. trav. bot. neérl.*, **25**, 1, 1928.
[13] Wey, v. d. H. G., *Proc. Kon. Akad. Wetensch., Amsterdam*, **34**, 874, 1931.

# THE BASIS OF THE DEPENDENCE OF VISUAL ACUITY ON ILLUMINATION

BY WARREN W. WILCOX

PSYCHOLOGICAL LABORATORY, UNIVERSITY OF KANSAS, LAWRENCE, KANSAS

Communicated December 8, 1931

It is a well-known fact that visual acuity, or the resolving power of the eye, varies with light intensity. Many speculations have been made regarding the cause of this variation. Particularly interesting is the theory recently put forward by Selig Hecht.[1] This theory not only attempts to provide a quantitative basis for the relation between visual acuity and illumination, but has certain implications which are of fundamental importance for sensory physiology in general.

Hecht assumes that the resolving power of the retina, like that of a photographic plate, must be determined by the "grain" of its surface, in other words, by the number of light-sensitive elements per unit area. "But the number of rods and cones in the retina is fixed anatomically. Therefore, it is necessary to suppose that the number of elements in the retina can be varied functionally." He supposes that the retinal elements vary in their absolute thresholds of response, only a small number being functional at the lowest intensities, and more and more coming progressively into function as the intensity is raised. On the basis of his theory regarding the nature of the photochemical system in the rods and cones, plus certain assumptions regarding the distribution of thresholds among the elements, he arrives at a quantitative deduction of the law relating acuity and intensity. This deduction agrees remarkably well with the empirical data of König on the visual acuity of the eye.

Hecht suggests that the number of active receptors furnishes the basis not only for visual acuity, but for the visual discrimination of intensities. The perception of a just noticeable difference in light intensity is determined by the addition or subtraction of a single functional receptor per unit retinal area. Number of elements, and not frequency of optic nerve discharge, governs the subjective differentiation of intensities. Carrying out an analysis of the data on intensity discrimination, and a comparison with the data on visual acuity, he finds quantitative confirmation of this additional hypothesis.

Houstoun[2] has lately developed a theory of visual acuity and of intensity discrimination which is essentially a modified form of the theory of Hecht.

In view of the far-reaching importance of Hecht's conceptions, it seemed desirable to make a careful study of the variations of visual acuity with illumination, with the purpose of obtaining, if possible, some direct experimental evidence regarding the underlying causes of these variations.

Although more than thirty researches have been devoted to the relation between acuity and illumination, these have yielded little more than a conflicting variety of empirical laws. No one, apparently, has attempted to make direct experimental tests which should furnish information regarding the causal factors involved.

The methods used by previous experimenters are unsatisfactory for such a purpose, since they involved so many sources of variation in addition to light intensity that a clear understanding of the intrinsic effect of intensity is impossible. Almost invariably, the acuity has been tested by varying the size of the retinal image of the test object as a whole, rather than merely the width of the "gap" which was to be resolved. Yet it is known that this change of size in itself affects the acuity. Moreover, no provision was made for separate illumination of the test object and its background; and an analysis of the individual effects of object intensity and background intensity could not be obtained.

In the present experiments, the situation was simplified to the utmost possible degree. The test object consisted of two narrow parallel vertical bars, of constant size but variable separation. The acuity was tested by determining the least interval between the bars which would yield a perception of doubleness. The bars and their background were separately illuminated. Monocular vision was employed, with careful control of fixation, the total area of stimulation being confined to the fovea. The field was viewed through an artificial pupil which was sufficiently small (4 sq. mm. in area) to eliminate the influence of changes in the size of the natural pupil upon the retinal illumination and upon the definition of the retinal image. Care was taken to ensure an adequate period of dark adaptation before each set of observations.

The test object was furnished by two narrow strips of polished metal, placed at 45° to the line of sight and illuminated from the side, thus functioning as mirrors. Behind these strips was a larger mirror, also placed at 45° and illuminated from the side, but by a separate source; this mirror served to furnish background illumination. The test object proper was a reduced virtual image of the two bars, 33 cm. from the eye. The image of each bar always subtended the same angular dimensions at the eye (width, 2′22″, and height, 20′). A rectangular diaphragm limited the size of the background field to 70′ width and 20′ height. Thus the observer saw a small rectangular field in the center of which were two narrow vertical stripes. The experimenter, by manipulating a fine screw, could alter the distance between the metal bars and thus test the observer's acuity.

The intensity reflected by these bars was controlled by varying the distance of the lamp illuminating them, by the use of ground glasses, and by the use of a sector disk. The background intensity was independently controlled by a similar method. The voltage across each lamp was main-

tained at exactly the same level throughout all the work. The physical quality of the lights was thus constant. There was no sensible color difference between the test object and background illuminations.

In the tests of acuity, the following procedure was adopted. Starting with the bars very close together, so that the observer perceived them as a single stripe, the experimenter slowly increased their separation, the observer reporting when a gap between them was first perceived. Under each set of conditions, a large number of observations was obtained, and the average visual angle subtended by the minimum perceptible interval was taken as the resolution threshold.

Although the test object employed has the advantage of great simplicity from the point of view of theory, yet the observations involved in the threshold determinations are very difficult. While in the case of such objects as Snellen hooks, Landolt rings, etc., the details are resolved rather suddenly, in this case the resolution is a gradual process, and the observer is perplexed to identify the exact moment at which he first sees two bars instead of one. The threshold is always high, as well as variable, for an untrained observer, but practice decreases it by several hundred per cent, and diminishes its variability. Eventually, the threshold reaches a stable, low level. The data described below are the minimum thresholds obtained after an extended period of practice. The results of the two observers who participated in the major part of the work, Mr. C. W. Clendenon and the writer, display, on the whole, a very satisfactory concordance.

The behavior of the resolution threshold was studied under a variety of conditions. In the first series of experiments, the background on which the two bars appeared was completely dark, and the intensity of the bars was varied throughout a wide range. In the second series, the bars themselves were dark, while the intensity in the background was varied. Still other series were made in which both test object and background were illuminated. Only the results of the first two series will be described here.

*Bright Bars on Dark Ground.*—The results of the first series, for one observer, are represented in figure 1, in which the resolution threshold in seconds of arc is plotted as a function of the logarithm of the test object intensity in photons. The white circles in this figure represent the threshold data. Each point stands for the average of a large number of observations—100 in most cases.

The significance of the other points and of the curve which also appear in this figure, as well as in the following one, will be explained later.

At low intensities, the threshold is very high. As the intensity is progressively increased, the threshold decreases, but only until a certain minimum is reached, after which it rises. Thus the prevailing assumption that increased intensity is always favorable to visual acuity is incorrect; under

certain conditions the acuity may actually diminish with an increase of intensity.

*Dark Bars on Bright Ground.* It is interesting to find that when the test object is dark, and the light in its background is varied, a different type of curve is obtained (Fig. 2). In this case, the threshold decreases continually with rising intensity, the rate of decrease becoming progressively less.

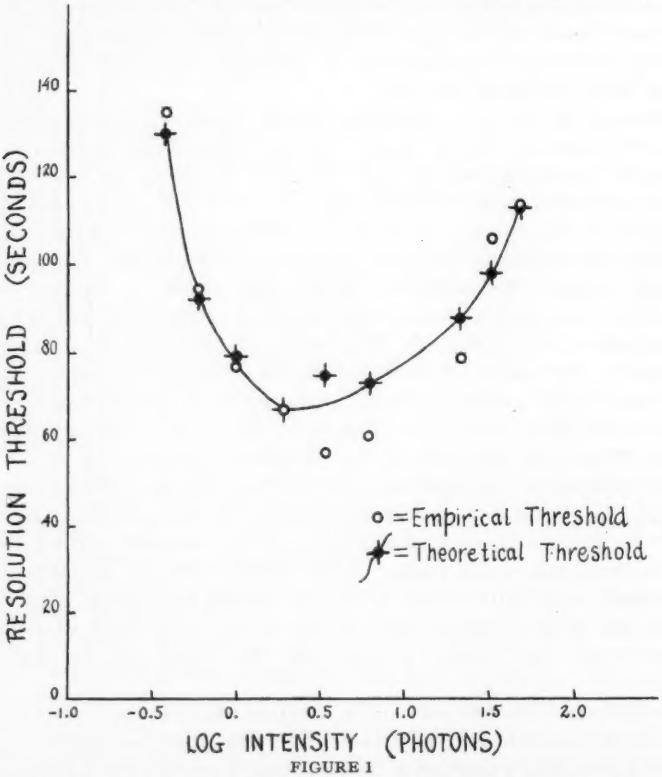



FIGURE 1

Showing the variation of the resolution threshold with intensity of the test object, in the case of bright bars on a dark background. Observer, Clendenon.

Thus the intensities in test object and background do not act symmetrically in regard to their effects on visual acuity. It follows from these results that it is impossible to state a single law expressing the relation between visual acuity and "intensity;" the acuity depends not only on the strength of the light, but also on the particular way in which this light is distributed between test object and ground.

What is the cause of the variation of acuity with illumination? During the course of the above experiments, an incidental observation was made which suggested a possible answer to this question. It was noticed that the two parallel bars, although physically constant in width, appeared to undergo pronounced variations in width as the intensity was altered. It is of course a familiar fact that a very bright object on a dark ground appears

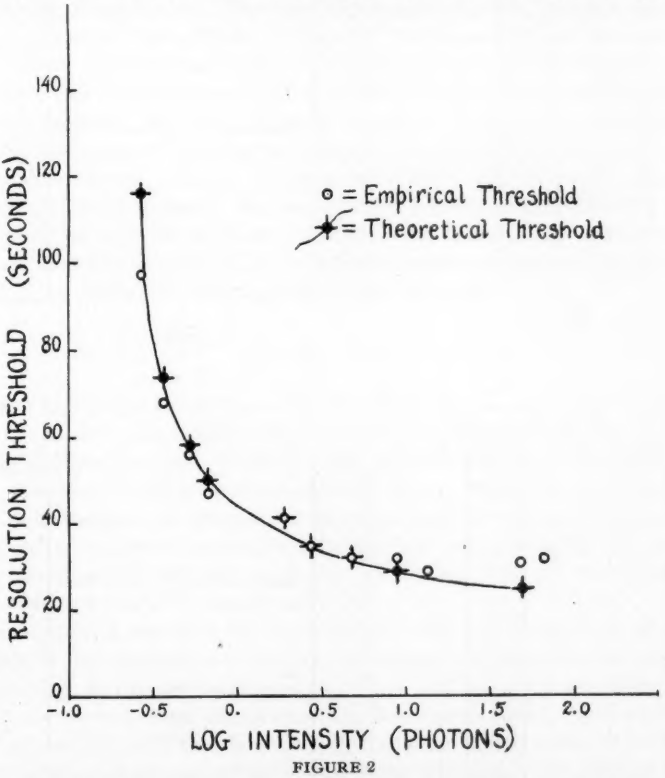
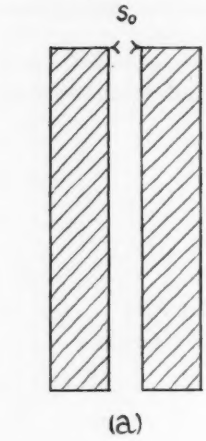



FIGURE 2

Showing the variation of the resolution threshold with intensity of the background, in the case of dark bars on a bright background. Observer, Clendenon.

subjectively enlarged; this is the phenomenon commonly described as "irradiation." It appeared likely that the rise of the threshold at high intensities in the first series (Fig. 1) was due to this factor. It was particularly interesting to note, however, that an apparent enlargement of the illuminated bars occurred at *low* as well as high intensities. Thus, if one began with very weak light, and gradually increased the intensity, the bars

appeared abnormally wide at first, then shrank continually, later expanding as the highest intensities were reached.



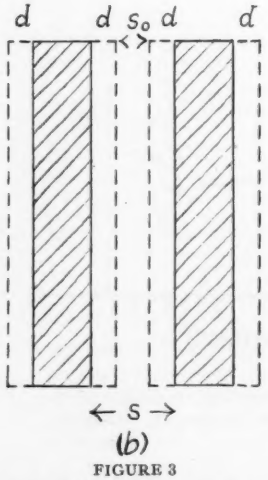
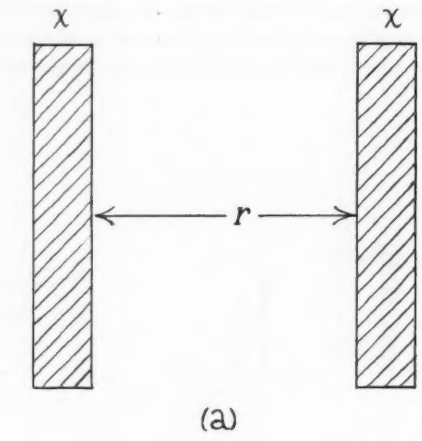




FIGURE 3

Illustrating the effect of irradiation on the resolution threshold. (Not drawn to scale.)



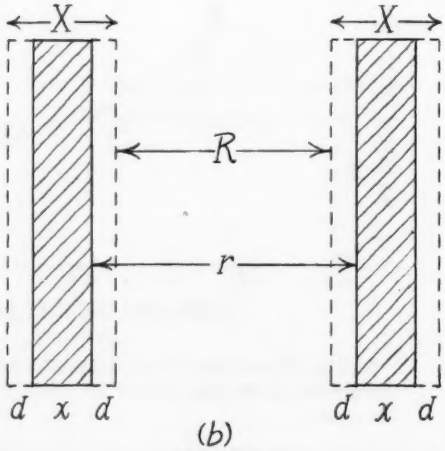




FIGURE 4

Illustrating the theory of the irradiation measurements. (Not drawn to scale.)

This fact suggested a working hypothesis, namely, that *these variations in apparent width are the sole basis of the changes in acuity with intensity.* If a

change of intensity causes the bars to spread (apparently) so that their inner contours are brought closer together, the bars must be separated by a correspondingly greater distance before the observer can perceive them as two discrete objects.

Figure 3 illustrates this hypothesis. In figure 3(a), the two rectangles drawn in solid lines represent the appearance which the bars would present in the absence of irradiation. The dotted contours in figure 3(b) represent the actual appearance when irradiation exists. We suppose that if irradiation were absent, the bars would have to be separated by a certain distance, $s_0$, before the gap between them could be perceived; this is the "ideal" threshold. If, however, the bars, originally placed at a distance $s_0$, apart, are increased in apparent width by irradiation, the gap between them will become imperceptible. If $d$ is the apparent displacement of each contour, then the distance between the two physical stimuli will have to be increased by an amount $2d$ before the separation of the apparent contours will be restored to the value $s_0$, and the gap once more perceived (Fig. 3(b)). Under these conditions, the threshold of resolution will be

$$s = s_0 + 2d.$$

Let us assume for the moment that the ideal value, $s_0$, of the threshold is known, i.e., the value which would be obtained if irradiation did not exist, so that the apparent sizes of objects would be proportional to their physical sizes, according to the laws of geometrical optics. Then it is obvious that if we have means of measuring $2d$, the relative shift of the adjoining contours, under any given intensity conditions, we can predict what the value of the threshold will be under these same conditions. We can thus subject our hypothesis to a strict quantitative test.

Fortunately, a measurement of irradiation could be obtained by a modification of the method devised by Volkmann.[3] Employing the same apparatus as in the threshold determinations, and adjusting the illumination to the desired value, the two vertical bars were gradually separated by the experimenter until the observer judged that the width of the interspace appeared equal to the sum of the widths of the two bars. The discrepancy between the physical and subjective widths then furnished a measure of the apparent shift of the contours.

Figure 4 illustrates the principle of this method. In figure 4(a), the two physical stimuli and their physical separation $r$ are represented, $x$ being the physical width of each bar. Figure 4(b) shows the appearance which the stimulus pattern presents to the observer. The bars are subjectively widened on each side by an amount $d$, so that their apparent width

$$X = x + 2d.$$

If $R$ is the *apparent* separation of the inner contours, then, according to the condition imposed,

$$R = 2X.$$

Furthermore,

$$R = r - 2d.$$

Combining these three relationships, we find

$$2d = \frac{r}{3} - \frac{2x}{3}.$$

Hence, if, after the observer has made the prescribed judgment, we measure the distance between the inner edges of the two physical stimuli, divide by 3, and subtract $^2/_3$ of the physical width of each stimulus bar, we obtain a numerical measure of the relative displacement of the two visible contours.

*Bright Bars on Dark Ground.*—In this case, as already noted, $d$ is large and positive at low intensities. As the intensity is raised, $d$ decreases to zero, and then increases once more.

*Dark Bars on Bright Ground.*—At the lowest intensities of the background on which the black bars appeared, the striking fact was noted that $d$ was *positive*, in other words that the bars were subjectively *widened*. Thus our measurements verified the existence of the so-called "negative irradiation" of Volkmann.[3] This apparent enlargement of the black bars was found to persist throughout a fairly wide range of intensities, although decreasing continually. Eventually an intensity was reached at which $d$ became zero. At higher intensities, the bars contracted still more, and $d$ became increasingly negative.

It will be seen at once that the results of both series of irradiation measurements are in harmony with our hypothesis regarding visual acuity.

Returning to our expression

$$s = s_0 + 2d,$$

we note that, even though $d$ can now be numerically specified under any given intensity conditions, we need to know the value of $s_0$ before we can test the quantitative validity of our hypothesis.

It will be recalled that $s_0$ is the resolution threshold corresponding to a condition in which irradiation, either "positive" or "negative" is absent, i.e., in which the quantity $d$ is zero. Now, as a matter of fact, the experimental data which have already been described yield a means of determining the value of $s_0$. For example, in the case of dark bars on a bright ground, we saw that $d$ was positive at low intensities and negative at high intensities, passing through the value zero at some particular intermediate value ($I_0$) of the background intensity. If, then, returning to the data on

visual acuity, we determine the value of the threshold for the resolution of the same dark bars *at this same intensity* ($I_0$) of the background, we obtain the numerical magnitude of $s_0$.

Fortunately, in every series of experiments it was found that $d$ assumed the value zero at some point in the intensity range. This made it possible in all cases to construct curves representing the theoretical values of the resolution threshold at all intensities, as calculated on the basis of the working hypothesis. One could then determine how well this hypothesis was confirmed by the threshold data obtained experimentally.

It was in this fashion that the curves drawn in figures 1 and 2 were obtained. The black-crossed circles in each figure represent the individual computations, derived from the irradiation measurements, to which the curve was fitted. Each such point is the mean of at least 20 measurements. In the same figures, as already mentioned, the empirical resolution thresholds are represented by white circles.

In view of the difficulty involved in the measurements, both of visual acuity and of irradiation, the agreement between theoretical and empirical values is surprisingly good. The theory finds still further confirmation in the results of the other series of experiments, in which both test object and background were illuminated.

In virtue of the outcome of these experiments, we may infer, with a high degree of probability, that irradiation, or the apparent shift of contours, is the only cause of the variations of visual acuity with intensity. In particular, one may conclude that the factor invoked in the theories of Hecht and Houstoun, namely, a variation in the number of functional retinal elements, is not required to explain the dependence of acuity upon intensity.

From this it follows that the conception regarding the basis of intensity discrimination which is advocated by these authors is without empirical foundation. For there is no evidence to support the idea that the individual rods or cones vary significantly in their sensitivity to light.

Our experiments point to the following general conclusion. If we can speak of a "space sense" as something distinct from the "light sense," then we can say that there is no ground for supposing that the acuity of the space sense changes with variations of intensity. The changes in the resolution threshold find their adequate explanation in the behavior of the light sense, as exemplified in the phenomenon of the apparent shift of contours.

It still remains to explain this phenomenon, and to elucidate the laws governing its relation to light intensity and light distribution. The writer, however, will defer the discussion of this question to a later and more extended paper which will appear elsewhere.

The present paper is based upon a thesis submitted in July, 1931, to the Department of Psychology and the Faculty of the Graduate School of the

University of Kansas, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer is indebted to Dr. D. McL. Purdy for suggestions and criticism.

[1] Hecht, These PROCEEDINGS, **13**, 1927 (569–574); *J. Gen. Physiol.*, Baltimore, **11**, 1928 (255–281).

[2] Houstoun, R. A., *Phil. Mag.*, London, **8**, 1929 (520–529); Houstoun, R. A., and Shearer, J. F., *Ibid.*, **10**, 1930 (433–450).

[3] Volkmann, A. W., *Physiologische Untersuchungen im Gebiete der Optik*, Leipzig, 1863.

---

## *QUANTUM DYNAMICAL CORRECTION FOR THE EQUATION OF STATE OF REAL GASES*

BY HENRY MARGENAU

SLOANE PHYSICS LABORATORY, YALE UNIVERSITY

Read before the Academy, Wednesday, November 18, 1931

Recent progress in the understanding of intermolecular forces has revived theoretical interest in the equation of state of real gases. Several investigations[2] make it appear very probable that the forces which cause the deviations from the perfect gas law are due to the interactions of rapid electronic motions within the molecules; and calculations of the second virial coefficient $B$, based on these concepts, have produced reasonable agreement with experimental findings. It was only in the case of very light gases, such as $H_2$ and He, that satisfactory agreement could not be obtained. Here the discrepancy proved to be such that for low temperatures the calculated values of B were much smaller (algebraically) than the observed data.

A possible explanation of this difficulty was first pointed out by London,[1] who suggested that the existence of zero point energy associated with the vibratory motion of molecules in quantized collision states might render the attractive Van der Waals' forces partially ineffective. However, a detailed quantitative investigation of this phenomenon and its effects upon the equation of state has not yet been presented. Such an investigation will be the object of this paper. We shall begin by assuming—in accord with the results of papers listed under 1)—that the forces between any 2 molecules have a potential energy $\epsilon$ which if plotted against the distance of separation $r$ of the two molecules, yields a curve similar to figure 1. If now these molecules are regarded as mass points and their mutual potential energy $\epsilon(r)$ is substituted in the Schroedinger equation, this will lead to solutions corresponding to discrete vibrational energy states $E_1 \ldots E_n$ of negative total energy, as well as to continuous positive states. It is not difficult to show that the relative spacing of the discrete states,

and the position of the lowest possible state $E_1$ depend on the depth of the minimum of $\epsilon$ and on the masses of the molecules. In particular, for light gases like $H_2$ and He the lowest energy state lies very high, indeed not much below the $O$-axis. Moreover, it is likely that for He there exists but this one vibrational state.[3] This situation implies that all pairs of molecules bound together by their potential energy, i.e., having a negative total energy, cannot exist in that part of phase space which corresponds to the shaded region unless they be in one of the states $E_1 \ldots E_n$, while there is no limitation on the states of molecules which have a positive total energy. In order to obtain a correct equation of state it is necessary to modify the usual calculations by excludng the inaccessible region of phase space, and considering instead the set of discrete levels $E_1 \ldots E_n$.

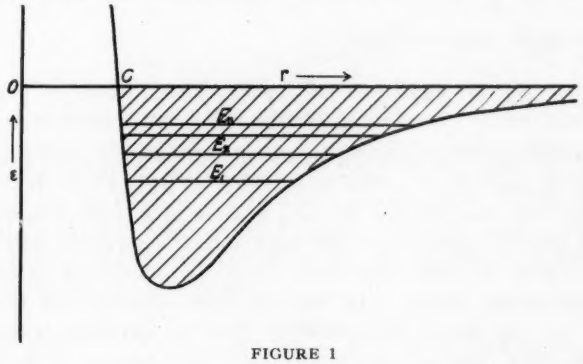



FIGURE 1

The pressure of a gas may be obtained from the thermodynamical relation

$$p = kT \frac{\partial}{\partial V} \log Z, \qquad (1)$$

where $Z$ is the complete partition function ("Zustandssumme" for the discrete states, phase integral for the continuous ones). If the gas consists of $N$ free mass points ($\epsilon = 0$)

$$Z = \frac{m^{3N}}{h^{3N}} \int \ldots \int e^{-\frac{m}{2kT}(v_1^2 + v_2^2 + \ldots v_N)^2} \, dv dV \qquad (2)$$

$dv$ is written for $dv_{1x} \ldots dv_{Nz}$, $dV$ for $dx_1 \ldots dz_N$. (2) is easily evaluated and leads to the law for ideal gases $pV = NkT$, if substituted in (1).

If the molecules, instead of being free, are considered to have a mutual potential energy $\epsilon(r)$ and the existence of quantized states is neglected, $Z$ differs from (2) only by the appearance of the total potential energy of the gas in the exponent of the integrand; that is,

$$Z = \frac{m^{3N}}{h^{3N}} \int \cdots \int e^{-\frac{1}{kT}\left[\frac{m}{2}(v_1^2 + \cdots v_N^2) + \epsilon(r_1 - r_2) + \cdots \epsilon(r_{N-1} - r_N)\right]} dv dV. \quad (3)$$

Here the integration extends over all velocity components from $-\infty$ to $+\infty$, and over all space. By $(r_i - r_j)$, the argument of $\epsilon$, we designate the relative distance between molecule $i$ and molecule $j$. A calculation of (3) may be found in textbooks on Statistical Mechanics.[4] The result is approximately

$$Z = \left(\frac{(2\pi mkT}{h^2}\right)^{3N/2} V^N (1 + \xi)^N \quad (4)$$

with $\xi = \frac{2\pi N}{V} \int_0^\infty (e^{-\frac{\epsilon(r)}{kT}} - 1) r^2 dr$. This expression, inserted in (1), yields the usual equation of state:

$$pV = NkT\left\{1 - \frac{2\pi N}{V} \int_0^\infty (e^{-\frac{\epsilon(r)}{kT}} - 1) r^2 dr\right\}. \quad (5)$$

It is evident that this equation ignores the effects of quantized collisions. A better approximation to the actual state of affairs may be obtained as follows: We consider the gas as a mixture of $n$ pairs of molecules in states of mutual quantization and neglect their interaction with similar pairs and with all other molecules. $N$ is the number of molecules in continuous energy states. The total partition function now consists of 3 parts: (1) the sum of state resulting from the vibratory motion of the pairs $\left(\sum_i e^{-\frac{E_i}{kT}}\right)^n$; (2) the phase integral due to the classical motion of translation of these pairs, which reduces to $\left[\frac{(4\pi mkT)^{3/2} V}{h^3}\right]^n$ and (3) $Z_p$, an integral like (3) but extended only over the accessible part of phase space.

$$Z = \left(\sum_i e^{-\frac{E_i}{kT}}\right)^n \left[\frac{(4\pi mkT)^{3/2}}{h^3} V\right]^n Z_p \quad (6)$$

Substitution in (1) shows that the first part of (6) contributes nothing to $p$, the second part only the partial pressure $\frac{nkT}{V}$, which may be neglected since, by ordinary statistical reasoning, $n$ may be shown to be very much smaller than $N$ at all temperatures of interest. The problem is then to evaluate $Z_p$.

The reasoning which leads from (3) to (4) involves the assumption that only "binary encounters" between the molecules can occur. This same restriction will here be imposed. The integrand of $Z_p$ may be transformed by the identity

$$\prod_{i=1}^{m} e^{a_i} = \Pi_m(e^{a_i} - 1) + \sum_{m-1}\Pi_{m-1}(e^{a_i} - 1) + \sum_{m-2}\Pi_{m-2}(e^{a_i} - 1) + \ldots + 1.$$

$\Pi_r$ indicates the product of $r$ terms $(e^{a_i} - 1)$ with different indices $i$ and selected from the total number of $m$ such terms, while $\Sigma_r$ denotes the sum over all possible products $\binom{m}{r}$ in number of $r$ terms. If now we write $\epsilon_i$ for $\epsilon(r_i - r_j)$, etc., allowing $i$ to take on values from 1 to $\frac{N(N-1)}{2} \equiv m$, and $W$ for the total kinetic energy of the $N$ molecules,

$$Z_p = \left(\frac{m}{h}\right)^{3N} \int_p \cdots \int e^{-\frac{W}{kT}} \left\{ \Pi_m\left(e^{-\frac{\epsilon_i}{kT}} - 1\right) + \sum_{m-1}\Pi_{m-1}\left(e^{-\frac{\epsilon_i}{kT}} - 1\right) + \ldots + \sum_1\left(e^{-\frac{\epsilon_i}{kT}} - 1\right) + 1 \right\} dv dV. \quad (7)$$

The limits of integration $p$ are defined as follows: If one molecule $\lambda$ is so far from all others that $\epsilon(r_\lambda - r_\mu)$ practically vanishes, where $\mu$ may be any other molecule of the assembly, then the velocity components of $\lambda$ are to be integrated from $-\infty$ to $+\infty$; however, if $\epsilon(r_\lambda - r_\mu)$ is different from zero the velocity components of $\lambda$ and $\mu$ run from values corresponding to a relative kinetic energy $\epsilon$ to $\pm\infty$. We illustrate the integration of (7) by considering a typical term:

$$I_r = \int_p \cdots \int e^{-\frac{W}{kT}} \sum_r \left(e^{-\frac{\epsilon(r_i - r_j)}{kT}} - 1\right) \cdots \left(e^{-\frac{\epsilon(r_k - r_l)}{kT}} - 1\right) dv dV.$$

There are altogether $r$ factors of the form $\left(e^{\frac{-\epsilon}{kT}} - 1\right)$. The assumption of binary encounters implies that if $\epsilon(r_i - r_j) \neq 0$, all other $\epsilon$'s involving either $r_i$ or $r_j$ must vanish. This in turn reduces to zero every term of $\Sigma_r$ in which any index occurs more than once and limits the number of terms of $\Sigma_r$ to $\frac{N!}{(N-2r)!\,r!\,2^r}$. These remaining terms contain the coördinates of $2r$ different molecules, so that the space integration over the other $N - 2r$ molecules simply gives $V^{N-2r}$, while the corresponding velocity integration over these molecules yields $g^{N-2r}$, where $g = \left(\frac{2\pi kT}{m}\right)^{3/2}$. The remainder of the integration over our typical term decomposes into successive independent integrations over the velocity and configuration spaces of pairs of molecules. Thus

$$I_r = \frac{N!}{(N-2r)!\,r!\,2^r} V^{N-2r} g^{N-2r} \eta^r,$$

where

$$\eta = \int_p \cdots \int e^{-\frac{m}{2kT}(v_1^2 + v_2^2)} \left[ e^{-\frac{\epsilon(r_1 - r_2)}{kT}} - 1 \right] dv_1 dv_2 dV_1 dV_2. \quad (8)$$

The evaluation of (7) by substitution of the various $I_r$ thus derived proceeds by a well-known analysis, for the details of which we wish to refer to footnote (4). If, for convenience, we define

$$\zeta = \frac{N}{2g^2V^2}\eta, \quad (9)$$

the result may be written in a form similar to (4):

$$Z_p = \left(\frac{2\pi mkT}{h^2}\right)^{3/2N} V^N (1 + \zeta)^N. \quad (10)$$

It is now necessary to calculate $\zeta$, defined by (9) and (8). The actual work involved in this process is somewhat lengthy and will not be written down in full. It is convenient to put $\eta = \eta_1 - \eta_2$, where $\eta_1$ is the result of an integration over all phase space, while $\eta_2$ is over the excluded domain of phase. Changing the variables in the volume integration to relative coördinates, $\eta_1$ is at once obtained:

$$\eta_1 = g^2V \int \left( e^{-\frac{\epsilon(r)}{kT}} - 1 \right) dV. \quad (11)$$

Using the abbreviations

$$\alpha^2 = \frac{m}{2kT}, \quad \rho^2 = \frac{2\epsilon}{m}, \quad \lambda = \sqrt{\rho^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2}, \quad \mu = \sqrt{\rho^2 - (z_2 - z_1)^2}$$

and writing $dV$ for an element of volume, as before, but $x$, $y$, $z$ for the components of velocity, the complete form of $\eta_2$ is seen to be

$$\eta_2 = \int\int dV_1 dV_2 \left( e^{\frac{-\epsilon(r_1 - r_2)}{kT}} - 1 \right) \int_{-\infty}^{\infty} e^{-\alpha^2 z_1^2} dz_1 \int_{z_1 - \rho}^{z_1 + \rho} e^{-\alpha^2 z_2^2} dz_2$$

$$\int_{-\infty}^{+\infty} e^{-\alpha^2 y_1^2} dy_1 \int_{y_1 - \mu}^{y_1 + \mu} e^{-\alpha^2 y_2^2} dy_2 \; X \int_{-\infty}^{\infty} e^{-\alpha^2 x_1^2} dx_1 \int_{x_1 - \lambda}^{x_1 + \lambda} e^{-\alpha^2 x_2^2} dx_2. \quad (12)$$

$\rho$ is to be considered as a function of the distance of separation, $r_1 - r_2$, of the two molecules. $\lambda^2$, $\mu^2$ and $\rho^2$ can never be greater than $\frac{2|\epsilon_{max}|}{m}$, so that the product of $\alpha^2$ and any of these three quantities has the maximum value $\frac{\epsilon_{max}}{kT}$. We shall develop (12) as a series of increasing powers of this parameter and thereby insure its rapid convergence for most cases of

physical interest. Only the evaluation of the last two integrals of (11) will be discussed as an example. Denoting their product by $J$ and changing variables,

$$J = \frac{1}{\alpha^2}\int_{-\infty}^{\infty} e^{-u^2}\,du \int_{u-\nu}^{u+\nu} e^{-v^2}\,dv, \quad \text{where } \nu = \alpha\lambda.$$

Now put

$$\int_0^x e^{-v^2}\,dv = G(x), \quad \text{so that}$$

$$J = \frac{1}{\alpha^2}\int_{-\infty}^{\infty} e^{-u^2}\,[G(u+\nu) - G(u-\nu)]\,du. \tag{13}$$

Expand

$$G(u+\nu) - G(u-\nu) = 2\left\{\nu G'(u) + \frac{\nu^3}{3!}G^{(3)}(u) + \frac{\nu^5}{5!}G^{(5)}(u) + \ldots\right\}.$$

The derivatives of $G$ are easily computed and inserted in (13). The integration is then to be performed only over terms like $u^n e^{-2u^2}$; it vanishes when $n$ is odd and otherwise yields $\frac{1.3\cdots(n-1)}{2^n}\sqrt{\frac{\pi}{2}}$. The result for $J$ is

$$J = \frac{2}{\alpha^2}\sqrt{\frac{\pi}{2}}\left(\nu - \frac{\nu^3}{6} + \frac{\nu^5}{40} + \ldots\right).$$

This may now be substituted in (11), and it must be remembered that $J$, through $\nu = \alpha\lambda$, is a function of the $y$'s and $z$'s. One continues by defining $\int_0^y J(x)e^{-x^2}\,dx = H(y)$ and developing the integral over $y_2$ in powers of $\alpha\mu$ in the manner just illustrated. After the integration over the velocity coördinates of (12) has been completed, the space integration over one of the $dV$'s may be carried out and there results

$$\eta_2 = g^2 V \int\left\{0.273\left(\frac{-\epsilon}{kT}\right)^{3/2} - 0.085\left(\frac{-\epsilon}{kT}\right)^{5/2} + 0.019\left(\frac{-\epsilon}{kT}\right)^{7/2} - 0.001\left(\frac{-\epsilon}{kT}\right)^{9/2} + \ldots\right\}\left(e^{-\frac{\epsilon}{kT}} - 1\right)dV.$$

The limits of $r$, the argument of $\epsilon$, are clearly the point $C$ in figure 1, at which $\epsilon$ becomes positive, and $+\infty$. Hence if we define a function $\varphi(x)$ to be 0 when $x < 0$, and equal to $0.273\,x^{3/2} - 0.085\,x^{5/2} + 0.019\,x^{7/2} - 0.001\,x^{9/2}$ for $x \geqq 0$,

$\eta_2 = 4\pi g^2 V \int_0^\infty \varphi\left(\frac{-\epsilon}{kT}\right)\left(e^{-\frac{\epsilon}{kT}} - 1\right) r^2 dr$, so that, recalling (9) and (11),

$$\zeta = \frac{2\pi N}{V} \int_0^\infty \left[1 - \varphi\left(\frac{-\epsilon}{kT}\right)\right]\left(e^{\frac{-\epsilon}{kT}} - 1\right) r^2 dr. \qquad (14)$$

We have mentioned that in most cases $Z_p$ (Eq. (10)), is the only quantity which needs to be retained in (6) if it is desired to derive an equation of state. Substituting (10) in (1), with $\zeta$ given by (14), the final form of our result becomes:

$$pV = NkT\left\{1 - \frac{2\pi N}{V} \int_0^\infty \left[1 - \varphi\left(\frac{-\epsilon}{kT}\right)\right]\left(e^{-\frac{\epsilon}{kT}} - 1\right) r^2 dr\right\}. \qquad (15)$$

It is needless to state that, like all similar classical results, (15) is valid only for small pressures, since in the last substitution $\zeta$ has been taken to be $\ll 1$, (log $(1 + \zeta) = \zeta$), and all through the analysis only binary encounters have been considered. (5) may be regarded as a first approximation to (15) much in the sense in which the perfect gas law approximates to (5). But the correction term in (15) is seen to become appreciable only at low temperatures, as a numerical application shows.[5]

Experimental observations on the second virial coefficient of He definitely call for this correction. However, perfect numerical agreement between experiment and theory has not yet been achieved.[5] Possibly this can be brought about only by a more basic fusion of the concepts of quantum dynamics and statistics in a manner that will not require the unrelated use of the Schroedinger equation for the purpose of obtaining $\epsilon$, and of Gibbs' statistics in the rest of the problem. This is a very fundamental and at present rather difficult issue; it involves among others the question as to the justification for treating the distance between two molecules, in the wave mechanical perturbation problem for determining $\epsilon$, as a parameter and not as an independent coördinate. As long as these points remain obscure, the method employed in this paper is probably the only available one.

[1] Eisenschitz and London, *Z. f. Phys.*, **60**, 491 (1930). London, *Ibid.*, **63**, 245 (1930). Slater and Kirkwood, *Phys. Rev.*, **37**, 682 (1931). Margenau, *Ibid.*, **37**, 1425 (1931).

[2] Margenau, *Phys. Rev.*, **36**, 1782 (1930). Wohl, *Z. f. Phys. Chem. B*, **14**, 36 (1931).

[3] Margenau, *Phys. Rev.*, **37**, 1014 (1931). The accuracy of the results there stated may now be improved by using a better approximation for $\Delta_2 E$; but they are qualitatively correct.

[4] Cf. for instance, R. H. Fowler, p. 168 et seq.

[5] Margenau, *Phys. Rev.*, **38**, 1785 (1931).

# NEW LINES IN THE K SERIES OF X-RAYS

BY WILLIAM DUANE

PHYSICAL RESEARCH LABORATORY, HARVARD UNIVERSITY

Communicated December 10, 1931

This article describes experiments with the K series of x-rays, chiefly those of molybdenum. The lines in the series are put in evidence by the Moseley photographic method. The distance from the slit in front of the target to the reflecting crystal equals that from the crystal to the photographic plate, thus corresponding to the Bragg focusing scheme. An accurate instrument with a worm gear causes the crystal to oscillate at constant speed back and forth over an angular range, which varies from five to sixty minutes of arc in different experiments. The oscillation is about a vertical axis of rotation. The distances from this axis of rotation to the slit in front of the target and to the center of the photographic plate amounts to 4725 mm., producing quite a large dispersion. The $\beta$ lines in the K series of Mo, shown in the accompanying diagrams, lay 0.88 mm. apart on the plate, corresponding to a difference between the wave-lengths of the $\beta_1$ and $\beta_2$ lines amounting to 0.00056 Ångström. The measurements by the ionization method of this doublet made some years ago by Allison and Armstrong in our laboratory and published in the *Physical Review* gave practically the same difference of wave-length between the $\beta_1$ and $\beta_2$ lines, namely, 0.000563 Ångström.

Plate 1 represents the K series lines of molybdenum, the reflecting crystal being calcite, and the 100 planes doing the reflecting. The horizontal ink lines below the spectrum lines represent the excursions produced by the oscillating crystal. The relative blackness of the various lines does not represent the relative intensities, for the times of exposure for the different lines were different from each other. The time of exposure of the $\alpha_1$ and $\alpha_2$ lines was 100 minutes. For the $\beta_1$ and $\beta_2$ lines it amounted to 150 minutes. For the lines of somewhat shorter wave-lengths, including the $\gamma$ line, the time of exposure amounted to $18^1/_2$ hours. This length of time makes the $\gamma$ line appear as black as the $\alpha_1$ line. The $\gamma$ line is, however, very much less intense than the $\alpha_1$ line.

A new line appears on this plate a little to the right of the $\gamma$ line, corresponding, therefore, to a shorter wave-length. Such a line in the K series of Mo has been reported by Leide, but, if the wave-lengths published by him in the *Comptes rendus* are correct, his line must lie further away from the $\gamma$ line than does the new line shown on Plate 1. As indicated on the other plates here presented, this new line is not a single line but a band.

The critical absorption of the K series of Mo, as carefully determined

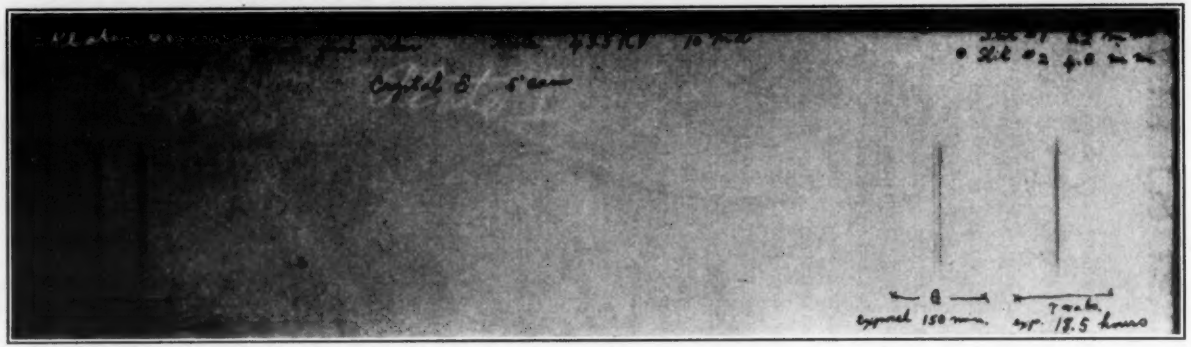



PLATE 1

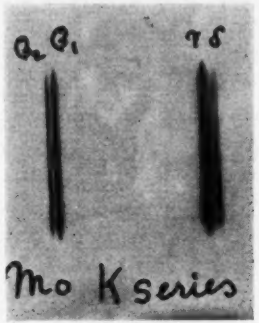



PLATE 2

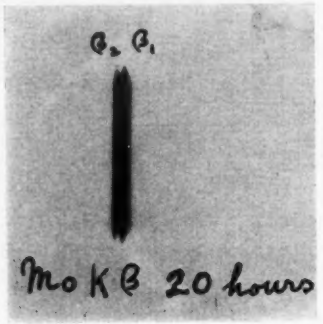



PLATE 3

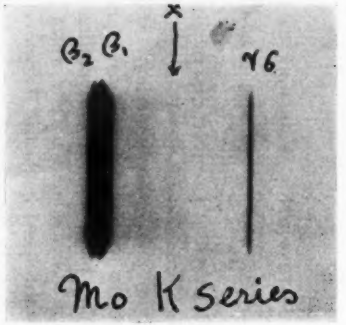



PLATE 4

in several ionization experiments performed in my laboratories, lies to the right of this new band, but very close to it. The new band, therefore, has wave-lengths that are only very slightly longer than that of the critical absorption.

As is well known, a number of interesting theories of x-radiation have been presented in recent years, such as the Bohr theory and amplifications of it. We have also corpuscular theories of radiation, such as those suggested during the first two or three years of this century and later put into such beautiful quantitative form by Einstein in his energy equation. If we use such a theory, we may regard the reflection of the x-rays by the crystal as representing transfers of momenta of radiation corpuscles to the crystal in quanta. We may also regard x-radiation from the point of view of the newer wave conception of de Broglie and the newer wave mechanics.

We now have a number of very interesting *physical* quantitative theories that have been proposed by various individuals. The importance of a physical theory depends (*a*) on the number of actual facts or phenomena that it clearly explains, (*b*) on the number of new laws and phenomena which it predicts and suggests for experimental verification and (*c*) on its clearness and the facility with which it can be used.

We may not, however, regard these theories as representing what, for lack of suitable English expressions, we may call real truth. They represent the ways in which certain types of human minds think, particularly if those minds have been trained in higher mathematics or along the lines of mathematical physics, as it is so well taught in Cambridge University, England. In the past these physical theories have not represented experimentally determined truth accurately and completely; so we must expect not them to in the future. They may represent nature in the sense of idealistic philosophy, an idea so clearly described by Bishop Berkeley in his analytical philosophy, namely, that the material world exists only as conceptions in the minds of human beings.

In this article we will use the theory that represents so many facts, namely, the theory that assumes x-rays to be the quantities of energy that are liberated when electrons fall in toward atomic nuclei.

I presented a preliminary report of the researches described in this article to the American Physical Society at its meeting in New York last February (1931). At that time I suggested that we might regard the new band mentioned above as due, at least in part, to the conductivity electrons falling into K positions made vacant by the electron bombardment of the target. I sent a copy of Plate 2 to Doctors DuMond and Hoyt, which they very kindly photometered for me, using the apparatus in the California Institute of Technology. They made, also, some interesting calculations, assuming that the entire band is due to conductivity elec-

trons distributed according to the Fermi probability laws, using Sommerfeldt's equation (*Physical Review*, Aug. 15, 1931, p. 839). These calculations indicated that the number of conductivity electrons in metallic molybdenum amounted to as many as 5 per atom and that there was a difference of potential amounting to as much as 15 volts. It may be that

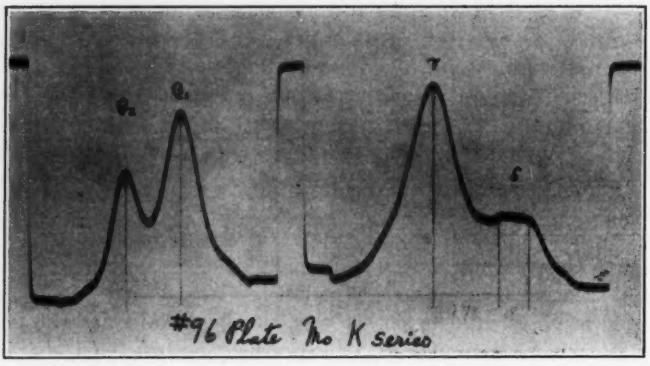
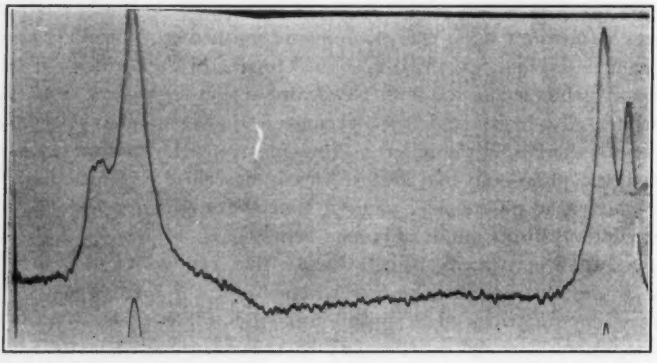



PLATE 5



$\delta$ $\gamma$ $\beta_1$ $\beta_2$

(The figures in plates 5 and 6 are in mutually reversed order.)

PLATE 6

the band should be regarded as representing in part a real $\delta$ line (O $\longrightarrow$ K electrons). This line would correspond to the small hump on the band next to the $\gamma$ line, and the rest of the band would correspond to the group of lines due to the conductivity electrons. This way of looking at it would

give a smaller difference of potential and a smaller number of conductivity electrons, perhaps 2 or 3 per atom. In order to decide these points of detail, we will have to wait until further analyses have been made of this radiation.

Plates 2, 3 and 4 represent the $\beta$, $\gamma$ and $\delta$ groups taken with much longer times of exposure. In plate 2 the exposure of the $\beta$ lines amounted to 10 hours and that of the $\gamma$ and $\delta$ lines to 91 hours. From the breadth of the $\gamma$ line in various plates one might assume that it is a doublet, which it should be according to certain theories. It looks, also, as if the $\beta_1$ line were more than a single line.

Plate 3 contains the $\beta$ lines produced by about 20 hours' exposure. $\beta_1$ is becoming quite fuzzy. There is, also, a very, very faint new line about half-way between the $\beta_1$ and $\gamma$ lines, which is so faint that probably it will be impossible to see it in the reproduction of this plate.

The line is marked $X$ in plate 4. In this plate the $\beta$ lines correspond to an exposure of $33^1/_2$ hours, whereas the $\delta$ and $\gamma$ lines correspond to only 10 hours' exposure. There is no known x-ray line belonging to any chemical element that could lie in its first-order reflection at the position marked $X$. In the second-order reflection one of the $\beta$ lines of cerium might lie at $X$. The wave-lengths of these cerium lines are not sufficiently well known to enable us to tell whether the $X$ line is a second-order reflection of cerium $\beta_1$, or not. If it is cerium $\beta_1$, it should be accompanied, theoretically, by another line close to it, cerium $\beta_2$. Further, if this is cerium $\beta_1$ the two $\alpha$ lines in the K series of cerium should appear in their second-order reflections close to the Mo K $\alpha$ lines. Experiments are in progress to test the radiation for the K $\alpha_1$ and $\alpha_2$ lines of cerium reflected in the second order. If they really exist they should be easily detected, and, if they are not detected, the $X$ line cannot be due to the radiation from cerium, but must be a new line.

No lines could appear at $X$ reflected in the 3rd or higher orders, for the constant voltage applied to the x-ray tube came from our storage battery and did not exceed 48. kv. in any experiment. This voltage would not produce lines short enough to be reflected to $X$ in the third order reflection.

The $\beta_1$ line of Mo on this plate appears to have a satellite on the short wave-length side of it. One would be inclined to think that the $\beta_1$ line is somewhat broader than $\beta_2$ on all of the plates, in spite of the fact that it is really twice as intense as $\beta_2$.

Plate 5 represents the $\beta$, $\gamma$ and $\delta$ lines of Plate 2 photometered by means of the photometer now existing in the Jefferson Laboratory. This instrument was designed by Professors Saunders and Crawford and very kindly loaned to me. The separation of the $\beta_1$ and $\beta_2$ lines does not appear to be as complete as on Plate 2 when examined by the eye. This, I understand,

is usually the case when close lines are photometered. The point is being studied by a new photometer, which has been constructed in our new physical research laboratory. It may be that the curves indicate a number of very close and very weak satellites.

Plate 6 represents the photometer curve kindly made for me at the California Institute of Technology. It corresponds almost exactly to the one obtained in the Jefferson Laboratory. The x-ray lines have been very kindly photometered for me in half a dozen laboratories throughout the United States. None of the curves, however, are superior to those represented in Plates 5 and 6.

A new $\delta$ line (O $\longrightarrow$ K electrons) has appeared on some photographs of the K series of tungsten made some time ago in my laboratory by Miss Armstrong, Assistant Professor of Physics at Wellesley College, when she was acting as one of my assistants. The spectra that she photographed also show two $\beta$ lines as well as the $\gamma$ and the two $\alpha$ lines of W.

Accurate details of these K series spectra of Mo and W will be published as soon as the photographs can be examined by means of the new photometer I have designed and am having set up in our new physical research laboratory.

It gives me great pleasure to sincerely thank my assistant, Mr. Lanza, who has spent so much time and care in setting up my apparatus and taking photographs.

---

## *NOTE ON PERRON'S SOLUTION OF THE DIRICHLET PROBLEM**

By Hassler Whitney†

Department of Mathematics, Princeton University

Communicated December 12, 1931

In 1923 Perron[1] gave a method of attacking the Dirichlet Problem, the only real difficulty of which lay in the proof of the fact that his function $u$ was harmonic. Simplifications of this proof were later given by several authors.[2] We give here a short proof involving only Harnack's first convergence theorem and inequality, both of which follow directly from the Poisson integral, and two lemmas from Perron's paper (Hilfssätze I and II). The convergence theorem is: *Let $u_1, u_2 \ldots$, be a sequence of functions harmonic in the closed region S, and converging uniformly in S to a limit u. Then u is harmonic in S.* The inequality we state in the form: *Let $\chi$ be harmonic and $\geqq 0$ within the circle K of radius a and with center P. Then if L is a circle about P of half the radius,*

$$\chi(Q) \leqq 3\chi(P)$$

*for any Q in L.*

Following Wiener[3] and Petrowsky,[3] we say a function $\psi$ is an upper function for $G + R$ and boundary values $f$ on $R$ if it is continuous and superharmonic ($M_K\psi \leqq \psi$ for any circle $K$) in $G$, and $\Psi \geqq \bar{f}$ on $R$. The solution of the Dirichlet problem is given by the function $u$, the lower limit of all upper functions. $u$ is easily seen to exist (see Perron's paper).

To show that $u$ is harmonic in $G$, let $P$ be any point of $G$, let $K$ be a circle about $P$ lying in $G$, and let $L$ be a circle about $P$ of half the radius. It is sufficient to show that to any $\epsilon > 0$ there corresponds a harmonic function $U$ such that

$$| U - u | \leqq \epsilon \text{ in } L.$$

For then to a sequence of $\epsilon$'s converging to 0 corresponds a sequence of harmonic functions approaching $u$ uniformly. Harnack's first convergence theorem now applies, and $u$ is harmonic in $L$. But as $P$ was any point of $G$, $u$ is harmonic in $G$.

Let $\psi$ be an upper function such that $\psi(P) \leqq u(P) + \epsilon/3$. As $M_K\psi \leqq \psi$ and $M_K\psi$ is an upper function, we have

$$u(P) \leqq M_K\psi(P) \leqq u(P) + \frac{\epsilon}{3}. \tag{1}$$

$M_K\psi$ is harmonic in $K$. I state that it is our required function; that is, for any $P'$ in $L$,

$$| M_K\psi(P') - u(P') | \leqq \epsilon. \tag{A}$$

If (A) is false, then there is a point $Q$ in $L$ for which

$$M_K\psi(Q) = u(Q) + \eta, \eta > \epsilon. \tag{B}$$

Let $\phi$ be an upper function for which $\phi(Q) < u(Q) + \eta - \epsilon$ Put $\phi' = \text{min. } (\phi, M_K\psi)$. Then $M_K\phi' \leqq \phi' \leqq \phi$, and hence

$$M_K\phi'(Q) < u(Q) + \eta - \epsilon.$$

This with (B) gives

$$M_K\phi'(Q) < M_K\psi(Q) - \epsilon.$$

But as $M_K\phi'$ is an upper function and $M_K\phi' \leqq \phi' \leqq M_k\psi$, $u \leqq M_k\phi' \leqq M_K\psi$ and (1) gives

$$M_K\psi(P) - M_K\phi'(P) \leq \frac{\epsilon}{3}.$$

Thus the function $\chi = M_K\psi - M_K\phi'$ is harmonic within $K$, is $\geqq 0$ there, is $\leqq \epsilon/3$ at $P$, and is $> \epsilon$ at $Q$ in $L$. But this contradicts Harnack's inequality, so (A) holds, and $u$ is harmonic in $G$.

The proof holds in any number of dimensions, if the constant in Harnack's inequality is changed properly.

* Presented to the American Mathematical Society, Dec. 28, 1931.

† NATIONAL RESEARCH FELLOW.

[1] O. Perron, *Math. Zeit.*, **18**, 42–54 (1923); we follow the notations in this paper.

[2] R. Remak, *Math. Zeit.*, **20**, 126–130 (1924); T. Radó and F. Riesz, *Ibid.* **22**, 41–44 (1925); R. Remak, *J. I. Math.*, **156**, 227–230 (1926). The proof here given is in essence like that of the last-named paper.

[3] I. Petrowsky, *Rec. Math. Moscou*, **35**, 105–110 (1928); N. Wiener, *J. Math. Phys. Mass. Inst. Techn.*, **4**, 21–32 (1925).

---

## *PROOF OF THE QUASI-ERGODIC HYPOTHESIS*

BY J. v. NEUMANN

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated December 10, 1931

1. The purpose of this note is to prove and to generalize the quasi-ergodic hypothesis of classical Hamiltonian dynamics[1] (or "ergodic hypothesis," as we shall say for brevity) with the aid of the reduction, recently discovered by Koopman,[2] of Hamiltonian systems to Hilbert space, and with the use of certain methods of ours closely connected with recent investigations of our own of the algebra of linear transformations in this space.[3] A precise statement of our results appears on page 79.

We shall employ the notation of Koopman's paper, with which we assume the reader to be familiar. The Hamiltonian system of $k$ degrees of freedom corresponding with the Hamiltonian function $H(q_1, \ldots, q_k, p_1, \ldots, p_k)$ defines a steady incompressible flow $P \longrightarrow P_t = S_tP$ in the space $\Phi$ of the variables $(q_1, \ldots, q_k, p_1, \ldots, p_k)$ or "phase-space," and a corresponding steady conservative flow of positive density $\rho$ in any invariant sub-space $\Omega \subset \Phi$ ($\Omega$ being, e.g., the set of points in $\Phi$ of equal energy). The Hilbert space $\mathfrak{H}$ consists of the class of measurable functions $f(P)$ having the finite Lebesgue integral $\int_\Omega |f|^2\rho d\omega$, the "inner product"[4] of any two of them $(f, g)$ and "length" $\|f\|$ being defined by the equations

$$(f, g) = \int_\Omega f\bar{g}\rho d\omega; \quad \|f\| = \sqrt{(f,f)}. \tag{1}$$

The transformation $U_t$ is defined as follows:

$$U_tf(P) = f(S_tP) = f(P_t); \tag{2}$$

obviously it has the group property

$$U_t U_s = U_{t+s}, \quad U_0 = I; \tag{3}$$

and in virtue of the conservative character of the flow, and the resulting invariance of $(U_t f, U_t g)$, it is unitary. The spectral reduction of $U_t$ in terms of its "canonical resolution of the identity" $E(\lambda)$[5] is furnished by a theorem due to Stone,[6] and gives us

$$U_t = \int_{-\infty}^{+\infty} e^{it\lambda}\, dE(\lambda), \tag{4}$$

this being the symbolic expression for the fact that, for all $f$, $g$ of $\mathfrak{H}$, we have, in terms of Stieltjès integrals,

$$(U_t f, g) = \int_{-\infty}^{+\infty} e^{it\lambda}\, d(E(\lambda)f, g). \tag{4'}$$

The pith of the idea in Koopman's method resides in the conception of the spectrum $E(\lambda)$ reflecting, in its structure, the properties of the dynamical system—more precisely, those properties of the system which are true "almost everywhere," in the sense of Lebesgue sets.

The possibility of applying Koopman's work to the proof of theorems like the ergodic theorem was suggested to me in a conversation with that author in the spring of 1930. In a conversation with A. Weil in the summer of 1931, a similar application was suggested, and I take this opportunity of thanking both mathematicians for the incentive which they furnished me for undertaking the investigations of this paper.

2. For the sake of brevity, we shall introduce the following notation:

We shall replace $\rho d\omega$ by $dv$, writing $\int_\Omega \rho d\omega = \int_\Omega dv$. By the "weight $\mu\Theta$ of the Lebesgue-measurable set $\Theta(\subset \Omega)$ with respect to the density $\rho$" will be meant the quantity $\mu\Theta = \int_\Theta \rho d\omega = \int_\Theta dv$. By a "zero set" we shall mean a set of zero weight, and hence, since throughout $\Omega$, $0 < \rho_1 < \rho < \rho_2$, a set of zero Lebesgue measure.

If $\Theta$ is a set of points $P$ of $\Omega$ or $\Phi$, we shall denote its characteristic function by $\chi_\Theta = \chi_\Theta(P)$; i.e., $\chi_\Theta(P) = 1$ or $0$ according as $\Theta$ does or does not contain $P$. If $f(P)$ is any measurable function, the set of points for which $f(P) > \lambda$, etc., will be denoted as usual by $[f(P) > \lambda]$, etc. We have the identity $[\chi_{[f>\lambda]} = 1] = [f > \lambda]$, etc.

By the strong convergence of a sequence $f_1, f_2, \ldots$ in $\mathfrak{H}$ $(f_n \longrightarrow f)$ will be meant that $\|f_n - f\| \longrightarrow 0$ as $n \longrightarrow \infty$. By weak convergence $(f_n \dot{\longrightarrow} f)$ we mean, on the other hand, that for an arbitrarily chosen $g$ of $\mathfrak{H}$, $(f_n, g) \longrightarrow (f, g)$ as $n \longrightarrow \infty$. It is shown that $f_n \longrightarrow f$ implies $f_n \dot{\longrightarrow} f$, but not conversely.[7] In general, expressions depending for their precise meaning on the nature of the convergence considered will be suffixed by the corresponding convergence symbol, thus we shall write

"separable ($\longrightarrow$)," "everywhere dense ($\dashrightarrow$)," etc. All these notions subsist if $n$ is replaced by one or more continuously varying parameters.

By $\mu$-convergence of a sequence of point sets $\Theta_1, \Theta_2, \ldots$ ($\subset \Omega$ or $\Phi$) will be meant the strong convergence of the corresponding measure functions: $\Theta_n \longrightarrow \Theta$ if $\chi_{\Theta n} \longrightarrow \chi_\Theta$, or, what is the same thing, $\mu[\Theta_n + \Theta - \Theta_n\Theta] \longrightarrow 0$ as $n \longrightarrow \infty$. Clearly $\Theta$, $\underline{\lim}\ \Theta_n$, and $\overline{\lim}\ \Theta_n$, will all differ by at most zero sets.

The greatest lower bound and least upper bound of a set [ ] will be denoted, as usual, by inf [ ] and sup [ ].

3. The starting point of our investigations is the construction of the operator

$$\sigma_{t,s} = \frac{1}{t-s}\int_s^t U_\tau d\tau \quad (s < t), \tag{5}$$

this being, as before, but the symbolic expression of the fact that for all $f, g$ in $\mathfrak{H}$,

$$(\sigma_{t,s} f, g) = \frac{1}{t-s}\int^t (U_\tau f, g)d\tau; \tag{5'}$$

the existence of $\sigma_{t,s}$ is easily proved.[8] We will show that, for each $f$ of $\mathfrak{H}$, $\sigma_{t,s}f$ is convergent ($\longrightarrow$) as $t - s \longrightarrow \infty$, irrespectively of the mode of variation of $s$, $t$.

We have from (5'):

$$\|\sigma_{t,s}f\|^2 = (\sigma_{t,s}f, \sigma_{t,s}f) = \frac{1}{t-s}\int_s^t (U_\tau f, \sigma_{t,s}f)d\tau$$
$$= \frac{1}{(t-s)^2}\int_s^t\int_s^t (U_\tau f, U_\sigma f)d\sigma d\tau;$$

since $U_\sigma$ is unitary and $U_\sigma^* = U_\sigma^{-1} = U_{-\sigma}$,[9]

$$\|\sigma_{t,s}f\|^2 = \frac{1}{(t-s)^2}\int_s^t\int_s^t (U_{\tau-\sigma}f, f)d\tau d\sigma,$$

which reduces, on making the change of variables $\tau - \sigma = x$, $\tau + \sigma = y$, to

$$\frac{1}{(t-s)^2}\int_{-(t-s)}^{+(t-s)}\ \int_{2s+|x|}^{2t-|x|} (U_x f, f)\frac{dx\,dy}{2}$$
$$= \frac{1}{(t-s)^2}\int_{-(t-s)}^{+(t-s)} (t - s - |x|)(U_x f, f)dx.$$

This may be calculated with the aid of (4'), inasmuch as the various changes in the order of integration are permissible, on account of the uniform convergence of the Stieltjès integral in (4')[10] (for all values of $t$—at present, $x$). Thus:

$$\| \sigma_{t,s} f \|^2 = \frac{1}{(t-s)^2} \int_{-(t-s)}^{+(t-s)} (t - s - |x|) \left[ \int_{-\infty}^{+\infty} e^{ix\lambda} \, d(E(\lambda)f, f) \right] dx$$

$$= \frac{1}{(t-s)^2} \int_{-\infty}^{+\infty} \left[ \int_{-(t-s)}^{+(t-s)} e^{ix\lambda} \, (t - s - |x|) dx \right] d(E(\lambda)f, f)$$

$$= \frac{2}{(t-s)^2} \int_{-\infty}^{+\infty} \left[ \int_{0}^{t-s} \cos(x\lambda) . (t - s - x) . dx \right] d(E(\lambda)f, f)$$

$$= \frac{2}{(t-s)^2} \int_{-\infty}^{+\infty} \frac{1 - \cos(t-s)\lambda}{\lambda^2} \, d(E(\lambda)f, f)$$

$$= \int_{-\infty}^{+\infty} \left[ \frac{\sin {}^1/_2(t-s)\lambda}{{}^1/_2(t-s)\lambda} \right]^2 d(E(\lambda)f, f).$$

This integral has a non-negative integrand and a non-decreasing expression after the $d$-sign; hence we may obtain an upper bound for it as follows: First, break it up into $\int_{-\epsilon}^{+\epsilon}$ and $\int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{\infty}$; ($\epsilon > 0$, to be considered later). Then replace the integrand in the first part by 1, and that in the second part by $\left[ \frac{1}{{}^1/_2(t-s)\epsilon} \right]^2 = \frac{4}{(t-s)^2\epsilon^2}$. Finally, replace the field of integration in the second by $\int_{-\infty}^{+\infty}$. We shall then have

$$\| \sigma_{t,s} f \|^2 \leq \int_{-\epsilon}^{+\epsilon} d(E(\lambda)f, f) + \frac{4}{(t-s)^2\epsilon^2} \int_{-\infty}^{+\infty} d(E(\lambda)f, f)$$

$$= \{ (E(\epsilon)f, f) - (E(-\epsilon)f, f) \} + \frac{4}{(t-s)^2\epsilon^2} (f, f).$$

Hence, as $t - s \longrightarrow +\infty$, $\overline{\lim} \| \sigma_{t,s} f \|^2 \leq (E(\epsilon)f, f) - (E(-\epsilon)f, f)$, and if, as $\epsilon \longrightarrow 0$ $\{E(\epsilon) - E(-\epsilon)\} f \longrightarrow 0$, we shall have, on letting $\epsilon \longrightarrow 0$, $\| \sigma_{t,s} f \|^2 \longrightarrow 0$, so that $\sigma_{t,s} f \longrightarrow 0$.

We now introduce the projection operator $E_0$ defined as follows: $(E(\epsilon) - E(-\epsilon))f \longrightarrow E_0 f$ as $\epsilon \longrightarrow 0$. The existence and projective nature of $E_0$ is easily deduced from the fact that $E(\epsilon) - E(-\epsilon)$ is a non-increasing "function" of $\epsilon$.[11] Thus, we are able to express the condition that $\sigma_{t,s} f \longrightarrow 0$ as $t - s \longrightarrow 0$ in the form: $E_0 f = 0$.

Suppose that $E_0 f = f$; then, since $E(\epsilon) - E(-\epsilon) \geqslant E_0$, we have $E(\epsilon)f = f$, $E(-\epsilon)f = 0$,[12] i.e., $E(\lambda)f = f$ for $\lambda > 0$, $= 0$ for $\lambda < 0$. Hence, for all $g$,

$$(U_t f, g) = \int_{-\infty}^{+\infty} e^{it\lambda} \, d(E(\lambda)f, g) = (f, g),$$

so that, for all values of $t$, $U_t f = f$.

Let $\mathfrak{M}$ be the linear manifold in $\mathfrak{H}$ corresponding with $E_0$. For every

$f$ of $\mathfrak{M}$, $E_0 f = f$; hence $U_t f = f$, and $\sigma_{t,s} f = f$, so that as $t - s \longrightarrow \infty$ we have

$$\sigma_{t,s} f \longrightarrow f = E_0 f. \tag{6}$$

For all $f$ orthogonal to $\mathfrak{M}$ on the other hand, we have

$$\sigma_{t,s} f \longrightarrow 0 = E_0 f. \tag{6'}$$

Now let $f$ be an arbitrary point of $\mathfrak{h}$; we can write $f = f_1 + f_2$, where $f_1$ is in $\mathfrak{M}$, and $f_2$, orthogonal to $\mathfrak{M}$. Then it will follow from (6), (6'), that we still have, as $t - s \longrightarrow 0$,

$$\sigma_{t,s} f \longrightarrow E_0 f. \tag{6''}$$

Throughout $\mathfrak{M}$, $U_t f = f$. Conversely, if $U_t f = f$, it will follow that $\sigma_{t,s} f = f$, and hence by (6''), $f = E_0 f$, i.e., $f$ belongs to $\mathfrak{M}$. Thus, $\mathfrak{M}$ is the class of all solutions of the equation $U_t f = f$ (i.e., the identity in $t$).

4. Let us examine $\mathfrak{M}$ more closely. Its elements $f$ are characterized by $U_t f = f$, and hence, in virtue of (2), by

$$f(P) \equiv f(P_t), \tag{7}$$

the $\equiv$ sign holding for all $t$ but with the possible exception of a zero set of points $P$ (in general, dependent on $t$). Hence, if $f$ is in $\mathfrak{M}$, $\mathfrak{F}(f)$ will be also, provided $\int |\mathfrak{F}(\lambda)|^2 d\lambda$ is finite.

Now $f$ can be expressed as the limit ( $\longrightarrow$ ) of functions of the form $\mathfrak{F}(f)$ where $\mathfrak{F}$ is susceptible of but a finite number of values, and these, in their turn, are linear combinations of similar functions susceptible only of the values 0 and 1.[13] The latter, being of the form $\mathfrak{F}(f)$, belong to $\mathfrak{M}$. If we denote by $\mathfrak{K}$ the class of all functions belonging to $\mathfrak{M}$ and taking on only the values 0 and 1, we may say that $\mathfrak{K}$ spans ( $\longrightarrow$ ) the closed ( $\longrightarrow$ ) linear manifold $\mathfrak{M}$.

If $f$ belongs to $\mathfrak{K}$, we shall write $[f(P) = 1] = \Lambda$, and $f = f_\Lambda (= \chi_\Lambda)$, (cf. § 2). Since $\mu\Lambda = \| f \|^2$, and $f$ is in $\mathfrak{H}$, $\mu\Lambda$ must be finite; and since $f$ is in $\mathfrak{K} \supset \mathfrak{M}$, it follows from (7) that the transformation $P \longrightarrow P_t$ changes $\Lambda$ by at most a zero set. These two properties, its finite weight and invariance under $P \longrightarrow P_t$, characterize $\Lambda$. We shall call any set having these properties a $\Lambda$-set. Evidently $\Omega$ will be a $\Lambda$-set if and only if $\mu\Omega$ is finite.

The class of all $\Lambda$-sets, being a subclass of the class of all measurable sub-sets of $\Omega$, is separable;[14] that is, there exists a sequence $\Lambda_1, \Lambda_2, \ldots$ of $\Lambda$-sets such that any $\Lambda$-set can be expressed as the limit (in the sense of $\mu$-convergence) of a subsequence of $\Lambda_1, \Lambda_2, \ldots$. By methods which we have used in another connection,[15] we are able to replace the sequence $\Lambda_1, \Lambda_2, \ldots$ by another sequence of $\Lambda$-sets, $\Lambda_1', \Lambda_2', \ldots,$

such that, firstly, any member of one sequence may be expressed in terms of elements of the other with the finite repetition of the operations of taking the logical sum $(+)$, the logical product $(\times)$ and the logical difference $(-)$; secondly, $\Lambda_1', \Lambda_2', \ldots$ may be set into one to one correspondence with certain rational numbers $\rho_n = \rho(\Lambda_n')$, $\Lambda_n' = \Lambda(\rho_n)$, $0 < \rho_n < 1$, in such a manner that $\rho_m < \rho_n$ implies $\Lambda(\rho_m) C \Lambda(\rho_n)$; and, thirdly, $\inf \rho_n = 0$ and $\sup \rho_n = 1$.

We now define the function $G(P)$ as follows:

$$G(P) \begin{cases} = \inf[\rho_n \text{ for which } P \text{ is in } \Lambda(\rho_n)]; \\ = 1, \text{ when no such } \rho_n \text{ exists.} \end{cases}$$

By its construction, $G(P)$ is invariant under $P \longrightarrow P_t$ (remaining unchanged apart from zero sets); and $\inf G(P) = 0$, $\sup G(P) = 1$. Since $[G(P) \leq \rho_n] = \Lambda(\rho_n)$, it follows that $f_{\Lambda_n'} = f_{\Lambda(\rho_n)} = F(G)$ (for we may set $F(\lambda) = \Lambda$ for $\lambda \leq \rho_n$, $= 0$ for $\lambda > \rho_n$); and therefore this property remains true for every $f_{\Lambda n}$,[16] and any $f_\Lambda$ of $\mathfrak{K}$ (cf. definition of $\longrightarrow$ for sets). And since every $f$ of $\mathfrak{M}$ is the limit $(\longrightarrow)$ of a sequence of linear combinations of functions of $\mathfrak{K}$, it follows that every such $f$ is a function of $G$.[17] Finally if $\lambda < \Lambda$, $\mu[G(P) \leqslant \lambda]$ is finite. For a $\rho_n > \lambda$ may be found, whereupon $[G(P) \leqslant \lambda] \subset [G(P) \leqslant \rho_n] = \Lambda(\rho_n) = \Lambda_n'$, and $\mu\Lambda_n' = \|f_{\Lambda_n'}\|^2$, which is finite for any $f_{\Lambda_n'}$ of $\mathfrak{K}(\subset \mathfrak{H})$.

Any function $G(P)$ like the above, such that $G(P_t) = G(P)$ for all $t$ except perhaps at zero-sets, such that $\lambda' = \inf G(P)$ and $\lambda'' = \sup G(P)$ exist, and that, if $\lambda < \lambda''$, $\mu[G(P) \leqslant \lambda]$ is finite, and which possesses the property that every $f$ of $\mathfrak{M}$ may be expressed as $\mathfrak{F}(G)$, shall be called a universal integral. We have shown that one universal integral always exists; obviously there are infinitely many.[18]

The class $\mathfrak{M}$ coincides with the totality of expressions $\mathfrak{F}(G)$ (with $\int_\Omega |\mathfrak{F}(G(P))|^2 dv$ finite). This makes it possible to express $E_0 f$ in terms of $G$. We shall give below, instead of our original method of computation, an abbreviated method for which we are grateful to Mr. M. H. Stone.

5. For an arbitrary $f$ of $\mathfrak{H}$, $E_0 f$ belongs to $\mathfrak{M}$, and is, accordingly, of the form $\mathfrak{F}(G)$. Let $\bar{\lambda} < \lambda''$ $(= \sup G)$, and define $\zeta(\lambda)$ to be 1 for $\lambda < \bar{\lambda}$ and 0 for $\lambda \leqslant \bar{\lambda}$. Let us set $\Lambda(\lambda) = [G(P) \leq \lambda]$ (we have seen that $\mu\Lambda(\lambda)$ is finite). Then we have, on the one hand,

$$\int_\Omega E_0 f(P) \zeta(G(P)) dv = \int_\Omega \mathfrak{F}(G(P)) \zeta(G(P)) dv^{20}$$

$$= \int_{\lambda'}^{\lambda''} \mathfrak{F}(\lambda) \zeta(\lambda) d\mu\Lambda(\lambda) = \int_{\lambda'}^{\bar{\lambda}} \mathfrak{F}(\lambda) d\mu\Lambda(\lambda),$$

and, on the other hand,

$$\int_\Omega E_0 f(P)\zeta(G(P))dv = (E_0 f, \zeta(G))$$
$$= (f, E_0\zeta(G))^{21} = (f, \zeta(G))$$
$$= \int_\Omega f(P)\zeta(G(P))dv = \int_{\Lambda(\bar{\lambda})} f(P)dv.$$

Thus:

$$\int_{\lambda'}^{\bar{\lambda}} \mathfrak{F}(\lambda)d\mu\Lambda(\lambda) = \int_{\Lambda(\bar{\lambda})} f(P)dv. \tag{8}$$

As $\lambda$ increases from $\lambda'$ to $\lambda''$, $\mu\Lambda(\lambda)$ goes in a non-decreasing fashion from 0 to $\mu\Omega$; and $\mu\Lambda(\lambda + 0) = \mu\Lambda(\lambda)$. In intervals $\lambda_1 \leq \lambda < \lambda_2$ where $\mu\Lambda(\lambda)$ is constant, $\Lambda(\lambda)$ changes at most by points $P$ of a zero set, i.e., $\lambda_1 < G(P) < \lambda_2$ can be true at most on a zero set; thus the behavior of $\mathfrak{F}(\lambda)$ in such intervals does not affect the relation $E_0 f = \mathfrak{F}(G)$—we may take $\mathfrak{F}(\lambda)$ constant upon them. It follows that the familiar theorems on the differentiation of Lebesgue integrals may be applied, the independent variable being here $x = \mu\Lambda(\lambda)$. From such considerations it follows that $\frac{d\left\{\int_{\Lambda(\lambda)} f(P)dv\right\}}{d\{\mu\Lambda(\lambda)\}}$ exists for all $\lambda$ in $\lambda' \leq \lambda < \lambda''$, except for a set of values of $\lambda$ for which the corresponding set of values $x = \mu\Lambda(\lambda)$ is of zero measure,[22] and this derivative is equal to $\mathfrak{F}(\lambda)$. The correspondence $P \longrightarrow x$, obtained by setting $\lambda = G(P)$, $x = \mu\Lambda(\lambda)$, carries a set $\Theta \subset \Omega$ into a set $\theta$ on the $x$-axis so that $\mu\Theta$ = measure of $\theta$.[23]

Thus we have, except for at most a zero set on $\Omega$,

$$E_0 f(P) = \left[\frac{d\left\{\int_{\Lambda(\lambda)} f(P)dv\right\}}{d\{\mu\Lambda(\lambda)\}}\right]_{\lambda = f(P)}. \tag{9}$$

This is naturally true for $\lambda = \lambda''$ only when $x = \mu\Lambda(\lambda'')$ exists, i.e., when $\mu\Omega$ is finite.

In the case $\lambda = G(P) = \lambda''$, we carry out the above process with $\zeta(\lambda) = 1$ for $\lambda = \lambda''$, $= 0$ for $\lambda \neq \lambda''$. On setting $\Lambda = [G(P) = \lambda'']$, the following becomes clear: If $\mu\Lambda = 0$, $\mathfrak{F}(\lambda'')$ does not affect $E_0 f(P)$. If $\mu\Lambda = \infty$, $E_0 f(P)$ must be zero; for it is constant on $\Lambda$, and belongs to $\mathfrak{H}$. If $\mu\Lambda$ is $> 0$ and finite, the considerations which lead to (8) show that

$$\mathfrak{F}(\lambda'')\cdot\mu\Lambda = \int_\Lambda f(P)dv. \tag{8'}$$

Since $\Lambda = \Lambda(\lambda'') - \Lambda(\lambda'' - 0) = \Omega - \Lambda(\lambda'' - 0)$, it follows that

$$E_0 f(P) = \frac{\int_{\Omega - \Lambda(\lambda'' - 0)} f(P)dv}{\mu[\Omega - \Lambda(\lambda'' - 0)]} \text{ (for } G(P) = \lambda''). \tag{9'}$$

This formula subsists when $\mu[\Omega - \Lambda(\lambda'' - 0)] = 0$ or $\infty$, provided we agree to replace the right-hand member by 0 in the case where the denominator vanishes (and hence the numerator also) or is infinite.

When $\mu\Omega$ is finite, (9′) leads to (9) (cf. [22]); otherwise, it forms its natural generalization.

6. Let $M$ and $N$ be any measurable sub sets of $\Omega$, with $\mu M$, $\mu N$ finite. Then $\chi_M(P)$ and $\chi_N(P)$ are in $\mathfrak{H}$, and we may apply our results to them. It follows from (5′) that $(\sigma_{t,s}\,\chi_N, \chi_M)$, which equals $\int_M \sigma_{t,s}\,\chi_N(P)dv$, is equal to

$$\frac{1}{t-s}\int_s^t \left\{\int_\Omega U_\tau \chi_N(P)\cdot\bar{\chi}_M(P)dv\right\}d\tau$$

$$= \frac{1}{t-s}\int_s^t \mu(S_\tau N \times M)d\tau \quad (S_t P = P_t, \text{ etc.})$$

$$= \int_m Z_{s,t}\,(P, N)dv,$$

where we have set $Z_{s,t}\,(P, N)$ equal to $\frac{1}{t-s}$ times the linear measure of the set of $\tau$-values for which $S_{-\tau}P\,(= P_{-\tau})$ is on $N$.[24] That is, $Z_{s,t}(P, N)$ is the mean time of sojourn of $P$ in $N$ between the times $s$ and $t$ (actually, with the sign of $\tau$ changed, but this is immaterial). Since the above is true for all $M$, we have

$$\sigma_{t,s}\chi_N(P) = Z_{s,t}(P, N), \tag{10}$$

and in virtue of (6″),

$$Z_{s,t}(P, N) \longrightarrow E_0\chi_N(P) = \chi_N^0(P) \text{ as } t - s \longrightarrow \infty, \tag{11}$$

in the sense of strong convergence in $\mathfrak{H}$.

On applying (9), (9′) to $f = \chi_N$, we have

$$\chi_N^0(P) = \left[\frac{d\{\mu[\Lambda(\lambda) \times N]\}}{d\{\mu\Lambda(\lambda)\}}\right]_{\lambda = x_n(P)} \tag{12}$$

when $G(P) < \lambda''$, and

$$X_n^0(P) = \frac{\mu[\Omega - N \times \Lambda(\lambda'' - 0)]}{\mu[\Omega - \Lambda(\lambda'' - 0)]} \tag{12′}$$

when $G(P) = \lambda''$. The right-hand members have a meaning except possibly for $P$ on a zero set: in (12), cf.;[22] in (12′), we take 0 when the denominator is infinite, or when it (and hence, the numerator) vanishes.

Let us express the content of (11) in the following three ways: (*A*) explicitly as strong convergence in $\mathfrak{H}$; (*B*) as point convergence of a sub sequence;[17] (*C*) as weak convergence, or rather as the implication of the latter regarding the inner product of (11) with an arbitrary $X_m$ ($\mu_m$ finite).

*A*. $\int [Z_{s,t}\,(P, N) - X_n^0(P)]^2 dv \longrightarrow 0$ as $t - s \longrightarrow +\infty$.

*B.* For every sequence $t_{1,s1}$; $t_2$, $s_2$; ... with $t_n - s_n \longrightarrow +\infty$, there is a sub sequence $t_{n_\nu}$, $s_{n_\nu}$ ($\nu = 1, 2, \ldots$) such that for all $P$ of $\Omega$ with the possible exception of a zero set, $Z_{s_{n_\nu}, t_{n_{q_\nu}}}(P, N) \longrightarrow X_n^0(P)$ as $\nu \longrightarrow \infty$.

*C.* For each subset $M$ of $\Omega$ of finite $\mu M$, $\int_M Z_{s,t}(P, N)dv \longrightarrow \int_M \chi_N^0(P)dv$ as $t - s \longrightarrow +\infty$.

We observe that although $\chi_N^0$ is expressed in (12), (12′) in terms of the non-uniquely determined universal integral $G$, its dependence upon $G$ is only apparent: each of (11), $A$, $B$ or $C$, determines $\chi_N^0$ uniquely.

7. The existence, for each point $P$, of the limit of the mean sojourn $Z_{s,t}(P, M)$ is a consequence of $A$, $B$ or $C$, and applies to any Hamiltonian system. Our system is *ergodic* if and only if this limit, $\chi_N^0(P)$, is independent of $P$, i.e., when

$$X_N^0(P) = C_n, \text{ a constant.} \tag{13}$$

When this is true, we must have in the case $\mu\Omega = \infty$ that $C_n = 0$; for otherwise, $\| \chi_N^0 \| = \infty$, whereas $\chi_N^0$ is in $\mathfrak{H}$. But when $\mu\Omega$ is finite, we have: $\int \chi_N^0(P)dv = (\chi_N^0, 1) = (E_0\chi_N, 1) = (\chi_N, E_0 1) = (\chi_N, 1) = \int \chi_N(P)dv = \mu N$. Hence, by (13),

$$C_n = \frac{\mu N}{\mu\Omega}. \tag{13′}$$

This is obviously true, from what was said earlier, when $\mu\Omega = \infty$.

It is now a simple matter to tell whether the system is ergodic or not, and we do not even need the more complete results of § § 4 and 5.

First, suppose that (13) (and consequently (13′)) is true for arbitrary $N$. Let $f$ be an element of $\mathfrak{M}$. Then, on the one hand, we have

$$\int \chi_N^0(P)\bar{f}(P)dv = \frac{\mu N}{\mu\Omega} \int_\Omega \bar{f}(P)dv = \bar{C}\mu \cdot N,$$

and, on the other hand,

$$\int_\Omega \chi_N^0(P)\bar{f}(P)dv = (\chi_N^0, f) = (E_0\chi_N, f) = (\chi_N, E_0 f) = (\chi_N, f)$$
$$= \int \chi_N(P)\bar{f}(P)dv = \int_N \bar{f}(P)dv.$$

From the equality of the final expressions for all $N$, we conclude that $f(P) = C$. Secondly, suppose that, conversely, every function of $\mathfrak{M}$ is a constant. Then $\chi_N^0$, belonging to $\mathfrak{M}$, is a constant, and (13) is true.

Thus the system will be ergodic if and only if $\mathfrak{M}$ consists exclusively of constants. This will be true if and only if $\mathfrak{R}$ consists exclusively of constants,[13] i.e., that the $\Lambda$-sets all differ from 0 or from $\Omega$ at most by zero sets. Thus we have proved the theorem:

*E. The system is ergodic if and only if every measurable set $\Lambda$ remaining invariant under $S_t$ (except for points of a zero set) reduces to 0 or to $\Omega$ (except for points of a zero set).*[25]

Since in $E$ the ergodic condition is the non-existence of any measurable $\Lambda$-sets ($\neq 0, \Omega$), one might be tempted to suppose that the ergodic condition as stated in (13) would have to hold for a correspondingly broad class of sets. This, however, is not the case: If (13) is true for all open sets $N$ of finite $\mu N$, it will be true, by continuity, for all $\mu$-limits of such sets (cf. §2),—i.e., for all measurable sets $N$ of finite $\mu N$.[26] Indeed, it is only necessary to require its truth for sets $N$ which are the sums of a finite number of the neighborhoods of an arbitrary "topologically equivalent system of neighborhoods" in $\Omega$,[26] for instance, for sums of finite numbers of spheres.

8. From a purely mathematical standpoint, the question as to the validity and most appropriate generalization of the ergodic hypothesis has been fully answered: these *special* problems have been reduced to the *general* problem of the integrals of the system—the structure of $G(P)$. Thus, the system is either ergodic, or else there is a non-constant $G(P)$, in which case $\Omega$ is decomposable into subsets like $[G(P) = \lambda_1]$, $[\lambda_1 \leq G(P) < \lambda_2]$, etc., upon which the flow has a sort of ergodic character, as is easily shown by means of (12), (12′).

But from the point of view of physics, there remains the difficult question as to the existence and nature of $G(P)$ in each particular case. It might happen that there are integrals of the system in the classical sense, i.e., analytic, or at least continuously differentiable, as would be true, for example, if $G(P)$ were of this character; in which case they could be used for the reduction of the dimensionality of $\Omega$ (cf.,[2] p. 315, last line). Or it might happen that no such integrals exist, in spite of the fact that $G(P)$ is non-constant. Conceivably this last situation is impossible when the Hamiltonian $H$ is analytic (or even continuously differentiable); if so, the proof of this fact would be most useful. But it appears that the proof could not be obtained alone from the general formal considerations in Koopman's method, i.e., from (3) and from

$$U_t F(f_1(P), f_2(P), \ldots) = F(U_t f_1(P), U_t f_2(P), \ldots), \qquad (14)$$

(cf.,[2] p 318). For (4), (14) remain true in the case that the one to one map $P \longrightarrow P_t$ of $\Omega$ upon itself is any one-parameter group of the following properties:

*a.* $P_t$ is a measurable function of $t$,

*b.* $P \longrightarrow P_t$ maps every measurable $P$-set in a measurable $P_t$-set with the same measure.

Since *a*, *b* permit of discontinuities of $P_t$, it is easy to give examples with only discontinuous integrals.

Indeed, even when $P \longrightarrow P_t$ is defined by means of equations of the type

$$\frac{\partial}{\partial t} x_1 = \alpha_1(x_1, \ldots, x_l), \ldots, \frac{\partial}{\partial t} x_l = \alpha_l(x_1, \ldots, x_l), \tag{16}$$

$(P:(x_1, \ldots, x_l))$, and when $b$ is valid—when $P \longrightarrow P_t$ is indeed an "incompressible continuous flow"—there are examples where all the integrals are discontinuous, and yet are not constants. (In an example, $l = 2$, $\alpha_1/\alpha_2$ is continuous in $P$, but $\alpha_1$, $\alpha_2$ themselves, discontinuous.)

We shall not pursue this question further.

We may observe, in conclusion, how remarkable it is that the concept of Lebesgue measure should play so important a rôle in a so essentially physical a question as the validity of the ergodic hypothesis, or, more generally, in the value of the limit of the mean sojourn, $\lim_{t-s \to +\infty} Z(P, N)$.

Even in the case where $N$ is an open set or, indeed, the sum of a finite number of spheres—which has an immediate physical significance—the function $\chi_N^0(P)$ given by the above limit does only need to be measurable! In the last analysis, one is always brought to the cardinal question: "Does $P$ belong to $\Theta$ or not?" where the set $\Theta$ is merely assumed to be measurable. The opinion is generally prevalent that from the point of view of empiricism such questions are meaningless, for example, when $\Theta$ is everywhere dense—for every measurement is of limited accuracy. The author believes, however, that this attitude must be abandoned, and gives the following reason as an argument:

Suppose that $\Omega$, in which $P$ varies, have a finite measure, $m\Omega$. Since $\Theta$ is measurable, it follows from a familiar theorem of Lebesgue that

$$\lim_{\epsilon \to 0} \frac{m[\Theta \times K(P, \epsilon)]}{mK(P, \epsilon)},$$

(where $K(P, \epsilon)$ is a sphere of center $P$ and radius $\epsilon$), exists at each point of $\Theta$, and $= 1$ with the exception of a zero set.[27] Similarly for all points of $\Omega - \Theta$, where it is zero, with the same exception. The same is true when the spheres are replaced by many other sorts of figures, e.g., cubes.[27] Consider a sequence of partitions of $\Omega$ into systems of disjoint cells, $Z_1^{(n)}, \ldots, Z_{k_n}^{(n)}$ $(n = 1, 2, \ldots)$, such that the maximum diameter $\epsilon_n$ of $Z_1^{(n)}, \ldots, Z_{k_n}^{(n)}$ approaches zero as $n \longrightarrow \infty$. The limited accuracy of measurements finds its expression in the fact that we have to consider different order of accuracy (viz., 1, 2, ...); where, by an experiment of order of accuracy $n$, shall be meant the mere process of distinguishing in which $Z_\nu^{(n)}$ $(\nu = 1, \ldots, k_n)$ $P$ lies.

Suppose that a measurement of order $n$ has established that, for in-

stance, $P$ lies in $Z_\nu^{(n)}$; then the (geometric) probability that $P$ belong to $\Theta$ is $\frac{m[Z_\nu^{(n)} \times \Theta]}{mZ_\nu^{(n)}}$. So that if

$$\frac{m[Z_\nu^{(n)} \times \Theta]}{mZ_\nu^{(n)}} < \delta \text{ or } > 1 - \delta \quad (\delta > 0), \tag{15}$$

we know with a probability $> 1 - \delta$ the answer to the question, "Is $P$ in $\Theta$ or not?" The fact that we will be able to answer this question with a probability $> 1 - \delta$ of being right has the *a priori* probability (i.e., before the observation is made) of

$$w_n^{(\delta)} = \frac{\sum_\nu'^{(n)} m Z_\nu^{(n)}}{\sum_{\nu=1}^{Kn} m Z_\nu^{(n)}} = \frac{\sum_\nu' m Z_\nu^{(n)}}{m\Omega},$$

where $\sum_\nu'^{(n)}$ represents the summation over all values of $\nu$ satisfying (15). If we could prove that $w_n^{(\delta)} \longrightarrow 1$ as $n \longrightarrow \infty$, it would become clear that, granted a sufficiently high accuracy of experiment, the above question could be answered with an arbitrarily great degree of certainty—i.e., the question has physical meaning. (This is seen by taking, e.g., $w_n^{(\delta)} > 1 - \delta$).

Suppose that $w_n^{(\delta)} \longrightarrow 1$ as $n \longrightarrow \infty$ is untrue. Then for infinitely many values of $n$, $w_n^{(\delta)} \leq 1 - \eta$ (for a certain $\eta > 0$); so that if $\sum_\nu''^{(n)}$ implies summation over all values of $v$ for which (15) is violated, we shall have

$$m \sum_\nu''^{,n)} Z_\nu^{(n)} \geqq \eta\, m\Omega.$$

The set $\Xi$ of all points $P$ which belong to infinitely many such sets $\sum_\nu''^{(n)} Z_\nu^{(n)}$ will then also have a measure $\geq \eta m\Omega > 0$, in virtue of a theorem of Arzela's.[28] If $P$ is on $\Xi$, it lies on infinitely many sets $\sum_{1\nu}''^{(n)} Z_\nu^{(n)}$; suppose it to be, for example, on $Z_{\nu_n}^{(n)}$. Since $\nu_n$ belongs, for infinitely many values of $n$, to $\sum_\nu''^{(n)}$, (15) is violated for these values, the ratio $\frac{m[\Theta \times Z_{\nu_n}^{(n)}]}{mZ_{\nu_n}^{(n)}}$ determines neither the limit 0 nor 1. But this is in contradiction with the theorem of Lebesgue, in the case where the $Z_\nu^{(n)}$'s are such that its hypothesis applies (e.g., when $Z_\nu^{(n)}$'s are cubes).

[1] For the formulation and critique of this theorem, cf., e.g., *Entykl. d. Math. Wiss.*, **4**, Art. 32 on Statistical Mechanics, by P. and T. Ehrenfest, specially 30–36. The original formulations are to be found in *Wien. Ber.*, **63**, [2] 679 (1871) (Boltzmann), and *Cambr. Phil. Soc. Trans.*, **12**, 547 (1879) (Maxwell).

[2] These PROCEEDINGS, **17**, [5] 315–318 (May, 1931).

[3] Cf., e.g., the discussion in the author's paper, "Allgemein Eigenwerttheorie Hermitescher Funktionaloperatoren" (*Math. Ann.*, **102**, [1] 108–111 (1929)). This paper, as well as the author's paper, "Zur Algebra der Funktionaloperatoren und Theorie der normalen Operatoren" (*Math. Ann.*, **102**, 3 (1929)), will be referred to in the present paper under the abbreviations E and A, respectively.

[4] Cf. E, 54–55, 109.

[5] Cf. E, 91–92.

[6] These PROCEEDINGS, **16**, [2] 172–175 (Feb., 1930); also, cf. a paper soon to appear in the *Ann. Math.*

[7] Cf., regarding these concepts, the article of Hellinger and Toeplitz in the *Math. Encyklopädie*, **2**, c. 13, 1435 (1928); further, cf. A, 378–381.

[8] Cf., e.g., the similar proof in E, 112, top.

[9] $R^*$ is the adjoint of $R$ in the terminology of matrices: the conjugate-transposed matrix. Cf., e.g., 112.

[10] The integrand, $e^{it\lambda}$, is uniformly bounded, the expression after the $d$-sign, $(E(\gamma), f)$, of bounded variation: $\int_{-\infty}^{+\infty} d(E(\gamma)f, f) = (f, f)$.

[11] E, 91, 77–78.

[12] Cf. the theory of projection operators outlined in E, 74–78; similarly for the discussion to follow.

[13] Cf. E, 110.

[14] Cf. Hansdorff, *Mengenlehre*, 127 (1927), line 6; or E, 110.

[15] Cf. E, 110.

[16] Cf. the corresponding construction in the proof of theorem 10; in A, 401–402. There, the permutable projections $E_1, E_2, \ldots$ and $F_1, F_2, \ldots$ took the place of the $\Lambda$-sets $\Lambda_1, \Lambda_2, \ldots$ and $\Lambda_1', \Lambda_2', \ldots$; but this distinction can be abolished by replacing each $\Lambda$-set by the operator, $E_\lambda$: $f(P) \longrightarrow \chi_\Lambda(P)f(P)$.

[17] For clearly, $\chi_{\Lambda+M} = \chi_\Lambda + \chi_M - \chi_\Lambda \cdot \chi_M$, $\chi_{\Lambda \times M} = \chi_\Lambda \chi_M$, $\chi_{\Lambda - M} = \chi_\Lambda - \chi_\Lambda \chi_M$.

[18] If the sequency $f_1, f_2, \ldots$ converges ($\longrightarrow$), a subsequency will converge in any point, excepted a 0-set (cf., f.i. E, 111). Therefore a limit of functions of $G$ is a function of $G$.

[19] Thus, e.g., each $T(G)$ is one, if $T(\gamma)$ is a monotonically increasing function.

[20] On the other hand, our construction shows that it suffices to confine $\mathfrak{F}$ to the second Baire class (the convergence is pointwise convergence except for zero sets).

[21] This transformation from Lebesgue to Stieltjès integrals goes back to Lebesgue. Cf. *Ann. de l'École Normale*, **3**, 27 (1910), p. 407. It is sufficient to establish it for a real variable, for it is then easily extended to an arbitrary $\Omega$, which may always be mapped in a measure-preserving manner upon the real axis. Maps of this sort are given by Lebesgue (loc. cit.) for $n$-dimensional space, and may easily be extended to $\Omega$.

[22] Since $\zeta(G)$ belongs to $\mathfrak{M}$, it is left unchanged by $E_0$.

[23] At points of discontinuity of $x = \mu\Lambda(\lambda)$, where $x$ experiences a jump of a whole interval, the differential quotient has a meaning, and is equal to the difference quotient between $\lambda + 0$ and $\lambda - 0$:

$$\frac{\int_{\Lambda(\lambda+0)-\Lambda(\lambda-0)} f(P)d\nu}{\mu[\Lambda(\lambda+0) - \Lambda(\lambda-0)]}.$$

[24] Cf. the author's paper, "Über Funktionen von Funktionaloperatoren," *Ann. Math.*, **32**, [2] 196 (1931), (Satz 3), as well as the reference (20).

[25] This transformation consists in a change in the order of integration; since in the Lebesgue integrals that appear, everything is bounded, it is permissible.

[26] For $\mu\Omega = \infty$, naturally only the former will come into question.

[27] Cf. E, 110.

[28] The generalization to other figures is to be found in its broadest form in Carathéodory, *Vorlesungen über reelle Funktionen* (Leipzig-Berlin, 1918), 492–494, in particular, Theorem 3. The function $f(P)$ appearing there is to be defined as $f(P) = X\Theta(P)$.

[29] Cf., e.g., de la Valleé Poussin, *Cours d'Analyse infinitésimale*, **1**, 2 (Louvain-Paris, 1909), pp. 68–69.

# SEVEN THEOREMS IN THE PROBLEM OF PLATEAU

BY JESSE DOUGLAS

DEPARTMENT OF MATHEMATICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Communicated November 24, 1931

The writer has recently completed the redaction of the manuscript of a paper presenting in complete form his results on the problem of Plateau for two contours; most of these results have been in my possession for some time past.[1] They are embodied in the form of seven theorems, which it is the purpose of this note to state. The complete paper will be published in the *Journal of Mathematics and Physics* of the Massachusetts I stitute of Technology.

The two contours $\Gamma_1$, $\Gamma_2$ are Jordan curves in euclidean space, of $n$ dimensions for theorems I, V, of two dimensions for theorem IV, and of three dimensions for theorems II, III, VI, VII. Always, $\Gamma_1$ *and* $\Gamma_2$ *are supposed not to intersect one another.*

With $\Gamma_1$, $\Gamma_2$ are associated three positive numbers, finite or $+\infty$:

$$m(\Gamma_1), \quad m(\Gamma_2), \quad m(\Gamma_1, \Gamma_2).$$

Concretely, these are, respectively, the least areas that can be bounded by $\Gamma_1$, by $\Gamma_2$, by $\Gamma_1$ and $\Gamma_2$;[2] but for our analysis they are the lower bounds of certain functionals

$$A(g_1), A(g_2), A(g_1, g_2; q),$$

where $g_1$, $g_2$ are arbitrary parametric representations of $\Gamma_1$, $\Gamma_2$ and $q$ is a parameter, $0 < q < 1$. Always, there is the inequality

$$m(\Gamma_1, \Gamma_2) \leqq m(\Gamma_1) + m(\Gamma_2),$$

or

$$e(\Gamma_1, \Gamma_2) = m(\Gamma_1) + m(\Gamma_2) - m(\Gamma_1, \Gamma_2) \geqq 0.$$

The functional of pairs of contours $e(\Gamma_1, \Gamma_2)$ is defined by the last formula for the case of finite $m(\Gamma_1, \Gamma_2)$. When $m(\Gamma_1, \Gamma_2) = +\infty$, we use the functional:

$$\bar{e}(\Gamma_1, \Gamma_2) = \limsup e(\Gamma_1', \Gamma_2') \geqq 0,$$

where $\Gamma_1'$, $\Gamma_2'$, contours with finite $m(\Gamma_1', \Gamma_2')$, (e.g., polygons), tend to $\Gamma_1$, $\Gamma_2$.

In theorems I, II, III, $m(\Gamma_1, \Gamma_2)$ is supposed finite, while in theorems V, VI, VII the contours are arbitrary Jordan curves, generally with $m(\Gamma_1, \Gamma_2) = +\infty$.

*Minimal surface* means one defined by the Weierstrass formulas:

$$x_i = R\ F_i(w),$$

$$\sum_{i=1}^{n} F_i^{12}(w) = 0.$$

THEOREM I. *Let $\Gamma_1$, $\Gamma_2$ be two Jordan curves not intersecting one another; let $m(\Gamma_1, \Gamma_2)$ be finite, and suppose we have the strict inequality:*

$$m(\Gamma_1, \Gamma_2) < m(\Gamma_1) + m(\Gamma_2),$$

*or*

$$e(\Gamma_1, \Gamma_2) > 0.$$

*Then there exists a doubly-connected minimal surface bounded by $\Gamma_1$, $\Gamma_2$. The area of this surface is $m(\Gamma_1, \Gamma_2)$.*

THEOREM II. *Let $\Gamma_1$, $\Gamma_2$ be two Jordan curves not intersecting one another, and let $m(\Gamma_1, \Gamma_2)$ be finite.*

*If the minimal surfaces $M_1$ and $M_2$, determined by $\Gamma_1$ and $\Gamma_2$ taken separately,[3] have in common a point that is regular for both of them $\left(\sum_{i=1}^{n} |F_i^1(w)|^2 > 0\right)$, then there exists a doubly-connected minimal surface bounded by $\Gamma_1$ and $\Gamma_2$.*

*The area of this surface is less than the sum of the areas of $M_1$ and $M_2$.*

THEOREM III. *Let $\Gamma_1$ and $\Gamma_2$, Jordan curves with finite $m(\Gamma_1, \Gamma_2)$, interlace. Then $\Gamma_1$, $\Gamma_2$ are the boundaries of a doubly-connected minimal surface.*

The writer's theory of the problem of Plateau includes the conformal mapping of plane regions as the special case $n = 2$. It is in this sense that the following theorem is to be understood.

THEOREM IV. *Let $\Gamma_1$, $\Gamma_2$ denote any two Jordan curves in the plane which enclose between them a region R. In the equations*

$$Z = g_1(z),\ Z = g_2(z),$$

*where Z and z denote complex variables, let Z describe $\Gamma_1$, $\Gamma_2$, respectively, when z describes two concentric circles $C_1$, $C_2$ of radii 1, $q$; $0 < q < 1$.*

*The range of values of the functional*

$$A(g_1, g_2;\ q) = \frac{1}{4\pi} \sum_{\alpha\beta} \int_{C_\alpha} \int_{C_\beta} |\ g_\alpha(z) - g_\beta(\zeta)\ |^2\ P(z, \zeta;\ q)dz\ d\zeta,^4$$

$\left(P(z, \zeta;\ q)\right.$ *being a certain elliptic function with periods* $2\pi$, $2\sqrt{-1}\log\frac{1}{q}\Big)$, *when all parametric representations $g_1$, $g_2$ of $\Gamma_1$, $\Gamma_2$ and all values of q are considered, will consist exactly of all positive real numbers $\geqq$ the inner area[5] of the region R. This minimum value will be attained for a certain (essentially*

*uniquely determined) parametric representation*

$$Z = g_1^*(z), Z = g_2^*(z),$$

*together with a unique value $q^*$ of $q$.*

*Then the integral formula of Cauchy:*

$$W = \frac{1}{2\pi i}\int_{C_1}\frac{g_1^*(z)\,dz}{z-w} + \frac{1}{2\pi i}\int_{C_2}\frac{g_2^*(z)\,dz}{z-w},$$ [4]

*defines a conformal transformation $w \longrightarrow W$ of the circular ring between $C_1$ and $C_2$ into the region $R$ between $\Gamma_1$ and $\Gamma_2$; this conformal transformation, furthermore, attaches continuously to the topological correspondence $g_1^*$, $g_2^*$ between the boundaries.*

In theorems V, VI, VII the restriction of finite $m(\Gamma_1, \Gamma_2)$ is removed from theorems I, II, III.

THEOREM V. *Any two Jordan curves $\Gamma_1$, $\Gamma_2$, not intersecting one another, for which*

$$\bar{e}(\Gamma_1, \Gamma_2) > 0,$$

*are the boundaries of a doubly-connected minimal surface.*

THEOREM VI. *If $\Gamma_1$, $\Gamma_2$ are any two Jordan curves not intersecting one another, and the minimal surfaces $M_1$ and $M_2$ determined by $\Gamma_1$ and $\Gamma_2$ separately*[3] *have a regular point in common, then there exists a doubly-connected minimal surface bounded by $\Gamma_1$, $\Gamma_2$.*

THEOREM VII. *Any two interlacing Jordan curves are the boundaries of a doubly-connected minimal surface.*

[1] "A General Formulation of the Problem of Plateau," presented to the American Mathematical Society, Oct. 26, 1929, abstract in *Bull. Am. Math. Soc.*, **36**, 50 (1930). "The Problem of Plateau for Two Contours," communicated to the same society, Sept. 27, 1930, abstract in the same publication, **36**, 797 (1930).

[2] The surfaces bounded by $\Gamma_1$ and $\Gamma_2$ separately are supposed to be simply-connected; those bounded by $\Gamma_1$, $\Gamma_2$ jointly, doubly-connected.

[3] The existence of the minimal surfaces $M_1$ and $M_2$ is assured by the writer's paper "Solution of the Problem of Plateau," *Trans. Amer. Math. Soc.*, **33**, 1, 263–321 (Jan., 1931), which gave the first general solution of the Plateau problem for a single contour.

[4] The sense of integration around $C_1$, $C_2$ is such that the circular ring between them is on the left.

[5] The upper bound of the area of a ring-shaped polygon whose boundaries $P_1$, $P_2$ encircle $\Gamma_2$ and are encircled by $\Gamma_1$; this is not always the same as the lower bound of the area of a ring-shaped polygon whose outer boundary $P_1$ encircles $\Gamma_1$ and whose inner boundary $P_2$ is encircled by $\Gamma_2$ (outer area).

# ON COMBINATORIAL TOPOLOGY

BY A. W. TUCKER

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated December 9, 1931

This note gives a preliminary account of some features of a paper on combinatorial topology to be published in more complete detail at a later date. The terminology and notation are patterned after that of Lefschetz,[1] but the cells $\{E_p\}$ with which we shall deal are of the abstract type considered by Mayer.[2] The advantage of using a general type of cell where the incidence numbers $\eta_{ij}^{p}$ may have any integral values, is described by Lefschetz (loc. cit., pp. 104–5). But there is one difficulty, viz., an $E_p$ cancelling out of the chain-boundary of an $E_{p+1}$ is not (combinatorially) distinguished from an $E_p$ not on the boundary of $E_{p+1}$ at all. This can be overcome by introducing a relation telling when an $E_p$ is *on the boundary of* an $E_q (p < q)$. The only connection of this qualitative relation with the quantitative one of incidence numbers is that if $E_p$ is not on the boundary of $E_{p+1}$ the incidence number is 0.

1. *Open and Closed Sub-Complexes.*—A complex $K$ is a set of cells such that the boundary-chain of a boundary-chain is always 0, i.e.,

$$F(F(C_p)) = 0$$

for an arbitrary chain $C_p$ on the set. If the cells of $K$ be separated into two sets $L$ and $J$ such that no cell of $J$ is on the boundary of a cell of $L$, examination of incidence matrices shows that $L$ and $J$ are complementary sub-complexes called *closed* and *open*, respectively. By suitable manipulation of incidence matrices the inter-relations of $K$, $L$ and $J$ may be read off. In particular if $K$ is "spherical" the identity of the homology characters of dimensions $p$ and $(p+1)$ of $L$ and $J$, respectively, may be obtained. (For a systematic treatment of incidence matrices it is found advantageous to adopt a canonical form in which the invariant factors appear in numerical order up the diagonal leading from the lower left corner, since then the incidence matrices of all dimensions of a complex may be simultaneously reduced to canonical form.)

2. *Blocks. Welds.*—A *p-block* of $K$ is defined to be a set of cells of $K$ such that if a cell belongs to the set so do all cells of dimensions $\geqq p$ of its closure on $K$ or $\leqq p$ of its star on $K$. (The *closure* on $K$ of a cell of $K$ consists of that cell and all cells of $K$ on its boundary. The *star* on $K$ of a cell of $K$ consists of that cell and all cells of $K$ on whose boundaries it lies.) By examining incidence matrices we find that a $p$-block of $K$ is a sub-complex of $K$. Indeed it is the most general type of sub-set of $K$ which can be directly shown to be a sub-complex from qualitative boundary

relations alone. A 0-block is a closed sub-complex, an $n$-block an open sub-complex.

Let the cells of $K$ be divided into complementary sets ${}^1K$, ${}^2K$. Under what conditions can ${}^2K$ be replaced by a cell $E_p$ so that ${}'K = {}^1K + E_p$ is a complex, every sub-complex of which has the same homology characters as the corresponding sub-complex of $K$, and moreover is closed (open) if the latter is closed (open)? The answer, found by manipulating incidence matrices, is that it is necessary and sufficient that ${}^2K$ be a $p$-cellblock of $K$, i.e., a $p$-block with the relative homology characters of a $p$-cell. Of course the incidence numbers of $E_p$ with cells of ${}^1K$ are determined by a choice of the basic relative $p$-cycle on ${}^2K$, and a cell of ${}^1K$ on the boundary of a cell of ${}^2K$ is supposed on the boundary of $E_p$ and conversely. We call the passage from $K$ to ${}'K$ a *$p$-weld*. It is a very general combinatorial operation. Subdivision is a special case of its inverse. (The concept of cellblock is a generalization of the idea, essentially due to Alexander, of "combinatorial cell"; cf. Lefschetz, loc. cit., p. 105.)

Analogous considerations show that a necessary and sufficient condition that ${}^2K$ may be dropped from $K$ without affecting homology characters, etc., is that ${}^2K$ be a block with the relative homology characters of the vacuous set.

3. *Multiplication.*—Comparison of products, joins and intersections leads to the conclusion that they can be obtained from a formal *$w$-multiplication*

$$E_p \overset{w}{\times} E_q = E_{p+q+w} \text{ or } 0$$

by taking $w = 0,1$ and $-n$, respectively. The single formula

$$F(C_p \overset{w}{\times} C_q) = F(C_p) \overset{w}{\times} C_q + (-1)^{p+w} C_p \overset{w}{\times} F(C_q) \qquad (3.1)$$

under suitable conventions yields the well-known formulae for the boundary-chains of products, joins and intersections.

This unification allows a theorem giving the homology characters of a product complex immediate application to joins and some intersections.

4. *Intersections.*—As indicated in §3 we introduce intersections as a special case of $w$-multiplication, viz., by taking $w = -n$. A fundamental assumption is that if $E_p$ is on the boundary of $E_q$, $E_p.E_r \neq 0$ implies $E_q.E_r \neq 0$. The intersection ${}^1K.{}^2K$ of complexes ${}^1K$ and ${}^2K$ is the set of non-vacuous cells ${}^1E_p.{}^2E_q$. From (3.1) it follows that ${}^1K.{}^2K$ is a complex.

We raise the question—is ${}^1K.{}^2K$ *identifiable* with ${}^aK (a = 1 \text{ or } 2)$, i.e., is it possible to weld the non-vacuous intersection cells in which

${}^aE_p$ figures into a $p$-cell ${}^{(a)}E_p$, for each ${}^aE_p$, so that the set $\{{}^{(a)}E_p\}$ forms a complex isomorphic with ${}^aK$? The answer is—a necessary and sufficient condition that ${}^1K.{}^2K$ be identifiable with ${}^aK$ is that for each ${}^aE_p$ the cells of ${}^bK$ ($b = 2$ or 1) with which ${}^aE_p$ has non-vacuous intersections form an $n$-cellblock whose basic relative $n$-cycle can be taken so as to coincide on the cellblock with a fixed $n$-chain of ${}^bK$.

5. *Manifolds.*—Consider the intersection complex $K.K^*$ of a complex $K$ and its *dual* complex $K^*$, the latter being defined abstractly in an obvious manner (cf. Mayer, loc. cit., p. 18). We assume that $E^i_p.E^{*j}_{n-p}$ only for $i = j$. Then, by the theorem stated at the end of § 4, $K.K^*$ is identifiable with $K^*$ if, and only if, $K$ has *star uniformity*, i.e., the star of each cell of $K$ is an $n$-cellblock whose basic relative $n$-cycle can be taken so as to coincide on the cellblock with a fixed $n$-chain of $K$. If in addition to star unformity $K$ has *closure uniformity* (= star uniformity for $K^*$, by definition) $K.K^*$ is identifiable with both $K$ and $K^*$, and so possesses absolute duality relations for Betti numbers and torsion coefficients,

$$R_p(K) = R_{n-p}(K)$$

$$\theta^p_i(K) = \theta^{n-p-1}_i(K)$$

Is such a $K$ an absolute manifold? Absolute duality relations alone do not justify calling it a manifold. One expects "invariance under subdivision" as well. Investigation of the latter leads to three results: (*a*) a necessary and sufficient condition that $K.K^*$ have star uniformity is that $K$ have star and closure uniformity; (*b*) a necessary and sufficient condition that $K.K^*$ have closure uniformity is that $K$ have *intercept uniformity*, i.e., that for all pairs of distinct cells $E_p,E_q$ of $K$ the common part of the star of $E_p$ on $K$ and the closure of $E_q$ on $K$ have the relative homology characters of the vacuous set; (*c*) a necessary and sufficient condition that $K.K^*$ have intercept uniformity is that $K$ have intercept uniformity—from which we conclude that star, closure and intercept uniformity make $K$ an absolute manifold.

By using "modular" star and closure uniformity apropos to $K.K^*$ taken modulo a suitable sub-complex, relative duality relations are obtained analogous to the Fundamental Duality Theorem of Lefschetz (loc. cit., p. 142). The investigation of "invariance under subdivision" in this case follows without much added complication the lines laid down in the last paragraph, and so contributes to the criteria determining a relative manifold. A supplementary contribution is furnished by a discussion of conditions on the modular cells, which indicates that a relative manifold is most tractable and informative when it has a "regular boundary."

It is interesting to note that the dual of the intersection complex $K.K^*$

can be considered as arising from $K$ and $K^*$ by a special 0-multiplication which could equally well have been used to develop the above manifold theory.

Of course between a manifold $M$ and its dual $M^*$ will automatically exist an intersection theory, viz., that given by $M.M^*$. But a theory for $M$ alone may also be constructed by first constructing a theory of looping coefficients for a "spherical" $M$ (cf. Lefschetz, loc. cit., p. 216). In the latter connection the result stated in §1 for a "spherical" $K$ enables us to obtain the analogue of Alexander's Duality Theorem.

6. *Elementary Subdivision.*—In dealing with joins (1-multiplication) it is convenient to introduce an $E_{-1}$ (suggested by the "1" of Alexander[3]) having an incidence number of 1 with each suitably oriented 0-cell and such that the join $E_pE_{-1} = E_p$ for an arbitrary $E_p$. If $E_{-1}$ may be added to a complex so as to get another complex the latter is called the *augmented* complex and the former is said to be *augmentable*. For us *elementary subdivision* means the inverse of a welding operation, which (inverse) is obtained by joining an augmented 0-cell to the boundary, either augmented or not, of the cell to be subdivided. A necessary and sufficient condition that a cell of a complex be elementarily subdivisible is that the cell have a closure, augmented or not as the case may be, with the homology characters of the vacuous set.

We call a complex *derivable* if all of its cells are elementarily subdivisible. The property of being derivable is invariant under elementary subdivision. By subdividing first the 1-cells, then the 2-cells, and so on, of a derivable $K$ we get a simplicial complex $K'$ which can always be geometrically realized. From our standpoint a (closed) simplicial complex may be defined as a set of cells such that (*a*) if $E_p$ is on the boundary of $E_q$ there exists a unique $E_{q-p-1}$ on the boundary of $E_q$ such that the join $E_pE_{q-p-1} = \pm E_q$, and (*b*) if $E_p$ is on the boundary of $E_q$ there exists an $E_r$, for each $p < r < q$, on the boundary of $E_q$ and on whose boundary $E_p$ lies. Such a complex has closure and intercept uniformity.

It turns out that any $K$ with closure uniformity is augmentable and derivable, and conversely. Intercept uniformity in $K$ ties up with star uniformity in $K'$. Hence our manifolds may be defined in terms of elementary subdivision.

[1] Lefschetz, "Topology," *Amer. Math. Soc. Colloquium Publications*, **12**, New York (1930).

[2] Mayer, "Über abstrakte Topologie," *Monatsh. f. Math. u. Phys.*, **36**, 1–42, 219–258 (1929).

[3] Alexander, "The Combinatorial Theory of Complexes," *Ann. Math.*, **31**, 294–322 (1930).

# LINES OF JULIA OF INTEGRAL FUNCTIONS[1]

BY W. SEIDEL[2] AND S. B. LITTAUER[3]

DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY

Communicated December 14, 1931

1. *Introduction.*—We present here the principal results of an investigation on the determination of lines of Julia of integral functions. Julia[4] established the fact that each integral function $f(z)$ possesses at least one ray (issuing from the origin) such that within every angle, however small, which the ray bisects, $f(z)$ assumes infinitely often each value with the possible exception of at most one. Such a ray is commonly called a *line of Julia*, denoted by *line J*. A number of writers have considered such half lines with regard to their properties, the degree of arbitrariness of the sets of such lines, and their location. We relate the existence of lines $J$ to asymptotic rays, to the order of the function on and in the neighborhood of a given ray, to the type of the function on the ray, and to the zeros of the function. Montel's[5] theory of normal families is used throughout.

2. *Sufficient Conditions.*—We establish the following theorems.

THEOREM I. *If the integral function* $w = f(z)$ *possesses two rays* $\theta = \theta_1$ *and* $\theta = \theta_2$, $(z = re^{i\theta}, 0 \leqslant \theta_1 < \theta_2 < 2\pi)$

*such that*

$$\lim_{r \to \infty} f(re^{i\theta_1}) = a$$

*and*

$$\lim_{r \to \infty} f(re^{i\theta_2}) = b \neq \infty,$$

*where* $a \neq b$, *then there exists at least one ray* $\theta = \phi$, $\theta_1 \leqslant \phi \leqslant \theta_2$ *which is a line J.*

It may happen that a given function assumes each value except perhaps one infinitely often in some angle of opening $\alpha$.[6] It has not been proved in general[7] that every such angle contains a line $J$. Under each of two additional restrictions this fact can be established. We have

THEOREM II. *Let the function* $f(z)$ *be analytic in an angle* $\theta_1 - \epsilon \leqslant \theta \leqslant \theta_2 + \epsilon$, $(0 \leqslant \theta_1 < \theta_2 < 2\pi, \epsilon > 0)$, *and in the angle* $\theta_1 \quad \theta \leqslant \theta_2$ *let it assume infinitely often every value with the possible exception of at most one. Furthermore, if either*

I. $\lim_{r \to \infty} f(re^{i\phi})$ *exists for some* $\phi$, $\theta_1 < \phi < \theta_2$; *or*

II. $f(z)$ *is bounded on both* $\theta = \theta_1$ *and* $\theta = \theta_2$, *then there exists at least one line J in the given angle.*

3. *Asymptotic Values.*—A ray $\theta = \theta_0$ on which a given integral function $w = f(z)$ is such that either

$$\lim_{r \to \infty} f(re^{i\theta_0}) = a,$$

$a$ finite, or for every positive $M$ there exists a positive constant $r(M)$ such that

$$|f(re^{i\theta_0})| > M$$

for $r > r(M)$, is called an asymptotic ray of $f(z)$.

It is well known that an asymptotic value of an integral function $w = f(z)$ corresponds to a non-algebraic singularity of the inverse function $z = \psi(w)$[8]. Further, two finite asymptotic values $a$ attained on different paths (1) and (2) are said to correspond to the same non-algebraic singularity of the inverse function if to each $\epsilon > 0$ there exists an $r(\epsilon) > 0$ such that for each pair of points $z_1 = r_1 e^{i\theta_1}$, $z_2 = r_2 e^{i\theta_2}$, where $r_1 > r(\epsilon)$ and $r_2 > r(\epsilon)$ there exists a path connecting $z_1$ and $z_2$, on which $|f(z) - a| < \epsilon$. If this is not the case, the asymptotic values are said to correspond to different non-algebraic singularities of the inverse function. The following theorems relate the existence of lines $J$ to the non-algebraic singularities of the inverse function.

THEOREM III. *If the integral function $w = f(z)$ possesses two rays $\theta = \theta_1$ and $\theta = \theta_2$, $(0 \leqslant \theta_1 < \theta_2 < 2\pi)$, such that*

$$\lim_{r \to \infty} f(re^{i\theta_1}) = a \neq \infty$$

*and*

$$\lim_{r \to \infty} f(re^{i\theta_2}) = a,$$

*then either $f(z)$ converges uniformly toward $a$ in the whole angle $\theta_1 \leqslant \theta \leqslant \theta_2$, and the two values $a$ correspond to the same non-algebraic singularity of the inverse function, or there exists at least one line $J$ in the given angle.*

The proof follows from an application of theorem I.[9] The following partial converse is also true.

THEOREM IV. *Under the hypotheses of theorem III, if $\theta = \phi_1$, $\theta_1 < \phi_1 < \theta_2$, and $\theta = \phi_2$, $\theta_2 < \phi_2 < \theta_1 + 2\pi$, are both lines $J$, then the two values $a$ attained asymptotically on $\phi_1$ and $\phi_2$ correspond to different non-algebraic singularities of the inverse function $z = \psi(w)$.*

By applying the preceding results one can obtain the following theorems.

THEOREM V. *A ray $\theta = \theta_1$, on which the integral function $w = f(z)$ attains no limit, but such that in every angular neighborhood containing it there exist asymptotic rays, is a line $J$.*

THEOREM VI. *An integral function $w = f(z)$ of finite order cannot converge on all rays to finite asymptotic values.*

4. *Order and Type.*—We establish the following condition based on the order of a function in an angle.

THEOREM VII. *If the integral function $w = f(z)$ possesses two rays $\theta = \theta_1$, and $\theta = \theta_2$, $(0 \leqslant \theta_1 < \theta_2 < 2\pi)$ such that*

$$|f(re^{i\theta_1})| = O(e^{r^\rho})$$

*and*

$$|f(re^{i\theta_2})| = O(e^{r^{\rho-\eta}}),$$

*where $\eta$ is some positive constant, then there exists at least one line $J$ on the boundary or in the interior of the angle $\theta_1 \leqslant \theta \leqslant \theta_2$.*

This follows directly from an application of a recent theorem of Mandelbrojt.[10]

We introduce the notion of type on a ray in the following manner.

*Definition* 1.—*Type.—An integral function $w = f(z)$ $(z = re^{i\theta})$ of order $\rho$ on a ray $\theta = \theta_0$ is said to be of type $\mu(\theta_0)$ if*

$$\mu(\theta_0) = \overline{\lim_{r \to \infty}} \log \frac{|f(re^{i\theta_0})|}{r^\rho}$$

*Definition* 2.—*Vanishing and Unbounded Types.—An integral function of order $\rho$ on a sequence of rays $\theta_n$ converging to $\theta = \theta_0$ is said to be of vanishing type on $\theta = \theta_0$ if*

$$\lim_{n \to \infty} \mu(\theta_n) = 0$$

*and of unbounded type if*

$$\lim_{n \to \infty} \mu(\theta_n) = \infty,$$

*where $\mu(\theta_n) > 0$ for all values of $n$.*

We find the following result on type.

THEOREM VIII. *A ray $\theta = \theta_0$ on which the integral function $w = f(z)$ is of finite order and vanishing or unbounded type is a line $J$.*

5. *Zeros of an Integral Function.*—It was shown by Gontcharoff[11] that if $\{a_n\}$ represents a sequence of zeros of an integral function clustering to a given ray, and such that $\left|\frac{a_{n+1}}{a_n}\right| < M$, where $M$ is a positive constant independent of $n$, and if the order of $f(z)$ is $\rho \leqslant 1/2$, the ray is a line $J$, while if the order of $f(z)$ is $\rho > 1/2$, there is a line $J$ within every angle of opening $2\pi - 2\pi/\rho$ containing the given ray. We obtain the following theorem on the zeros.

THEOREM IX. *Let $w = f(z)$ be an integral function and let $\{a_n\}$, $(n = 1, 2, \ldots)$ denote a sequence of zeros of $f(z)$, such that $n^\rho < |a_n| < n^{\rho+1-\epsilon}$, $\rho > 0$, $\epsilon > 0$, for all but a finite number of zeros; then, (I) if $\{a_n\}$ be confined*

*to any angle with vertex at the origin, there is a line J in the given angle, or in it f(z) converges uniformly to zero; (II) if $\{a_n\}$ clusters to a given ray, the ray is a line J, or on it f(z) converges to zero.*

[1] Presented to American Mathematical Society, October 31, 1931.

[2] NATIONAL RESEARCH FELLOW.

[3] NATIONAL RESEARCH FELLOW.

[4] G. Julia, *Leçons sur les fonctions uniformes à point singulier essentiel isolé*, Paris, chap. V, 102–118 (1924).

[5] P. Montel, *Leçons sur les familles normales des fonctions analytiques*, Paris (1927).

[6] See L. Bieberbach, *Math. Zeit.*, **3**, 175–190 (1919).

[7] See, however, G. Valiron, *Acta Mathematica*, **52**, 67–92 (1928), and H. Milloux, *Ibid.*, **52**, 189–255 (1929).

[8] See F. Iversen, *Compt. Rend.* (*Paris*), **166**, 156 (1918).

[9] See E. Lindelöf, *Acta. Soc. Fenn.*, **35**, Nr. 7 (1908).

[10] See S. Mandelbrojt, *J. Math.* (*Liouville*), **18**, 176 (1929).

[11] See Gontcharoff, *Compt. Rend.* (*Paris*), **185**, 1100–1102 (1927).

---

## *ON THE TIME AVERAGE THEOREM IN DYNAMICS*

BY EBERHARD HOPF*

HARVARD COLLEGE OBSERVATORY

Communicated December 4, 1931

1.—*Introductory.*—We consider a conservative dynamical system with an analytic Hamiltonian function $H(p_1, \ldots p_n, q_1, \ldots q_n)$. The phases $p_1, \ldots p_n, q_1, \ldots q_n$ of the system may be regarded as points $P$ in the phase space. After elapse of the time $t$ a point $P = P_0$ will be carried into a new position $P_t = T_t(P)$. Apart from singularities, $T_t(P)$ is well known to represent an analytic one parameter group of analytic one to one transformations of the phase space into itself,

$$T_s T_t = T_{s+t}, \quad (P_t)_s = P_{t+s}.$$

The phase volume is invariant under $T_t$. As $H = \text{const.}$ is an integral of the equations of motion, $T_t$ implies a group on the $(2n-1)$-dimensional manifold $H = c$. Similarly, we obtain a group on the $(2n - m - 1)$-dimensional manifold $H = c$, $H_\nu = c_\nu$, if the equations $H_\nu = c_\nu$, $\nu = 1_1 \ldots m_1$ represent $m$ further known integrals. In any case, a positive integral $m = \int \rho d\sigma$, $\rho > 0$, invariant under $T_t$, is associated with the considered manifold. We suppose that the group is free of singularities, i.e., that $T_t$ is analytically determined for all values of $t$. This situation occurs in well-known cases, for instance, in the case of the regularized problem of three bodies.

Generally, let $\Omega$ be an analytic manifold, and let $P \longrightarrow P_t$ $(P_s)_t = P_{s+t}$ be an analytic group of analytic one to one transformations of $\Omega$ into itself. A measure $m$ in the sense of Lebesgue is supposed to exist on $\Omega$ and to be invariant under those transformations. The well-known notation

$$(f, g) = \int_\Omega f(P)\overline{g(P)}dm, \quad \|f\| = \sqrt{(f, f)}$$

is used. A function $f(P)$ with finite $\|f\|$ represents a point in the Hilbert space, and the transition from the function $f(P)$ to the function $f(P_t)$ represents a linear and unitary transformation $U_t(f)$ of the Hilbert space into itself,

$$U_t(af + bg) = aU_t(f) + bU_t(g); \quad (U_t f, U_t g) = (f, g), \qquad (1)$$

the unitarity being due to the invariance of the measure $m$. The $U_t(f)$ form a linear one parameter group of unitary transformations,

$$U_t U_s = U_{s+t}. \qquad (2)$$

This connection between the motions of a dynamical system and linear groups of unitary transformations in the Hilbert space has been recently pointed out by B. O. Koopman.[1]

In this paper the author gives a simple and elementary proof of a remarkable theorem recently proved by J. v. Neumann in connection with the above correspondence principle.[2] In the dynamical case this theorem concerns the existence of the "time average" $f^*(P)$ of a function $f(P)$,

$$f^*(P) = \lim_{T=\infty} i.m. \frac{1}{T}\int_0^T f(P_t)\,dt,$$

in the sense of convergence in the mean on $\Omega$. v. Neumann's general theorem runs as follows:

THEOREM. *For any given point $f$ of the Hilbert space an "average point" $f^*$ is associated with the set of points $U_t(f)$, $-\infty < t < +\infty$, in such a way that*

$$\lim_{T=\infty} \left\| \frac{1}{T}\int_\alpha^{\alpha+T} U_t f dt - f^* \right\| = 0.$$

If, in the dynamical case, the function $f(P)$ is the characteristic function of a measurable point set $A$, $f = 1$ on $A$, $f = 0$ on $\Omega - A$, $f^*$ represents the mean time in which the wandering point $P_t$ crosses $A$. Professor Birkhoff discovered quite recently the remarkable fact that, in the case where $A$ is an open point set, $f^*$ is not only the limit in the mean, but the actual limit apart from a point set of zero measure.[3] The existence of

this mean time of crossing establishes the actual proof of a fundamental hypothesis of classical statistical mechanics.

v. Neumann's proof of the above theorem is based on the spectral representation of the operators $U_t$,

$$(U_t f, g) = \int_{-\infty}^{+\infty} e^{i\lambda t}\, d(E_\lambda f, g),$$

due to M. Stone. The spectral form is of some interest in connection with dynamics, since it opens a way toward the Fourier analysis of the general motion. It throws also a sidelight on convergence questions in dynamics and celestial mechanics. In general, the spectrum of a dynamical system will be a continuous one, i.e., the function $\varphi(\lambda) = (E_\lambda f, g)$ will not be a pure step function. This shows that in general the motions of dynamical systems cannot be approximated by trigonometric sums in any sense of uniformity within $-\infty < t < +\infty$. However, as far as concerns the proof of v. Neumann's above theorem, the usage of the spectral theory is avoidable.

2. *Proof of the Theorem.*—We start with two preliminary remarks due to v. Neumann (loc. cit.) For any point $f$ of the Hilbert space the integral $\int_a^b U_t f dt$ is a well-defined point of that space. Furthermore, we have

$$\left(\int_a^b U_t f dt,\ g\right) = \int_a^b (U_t f, g)\, dt.$$

We set

$$\varphi_{T,a} = \frac{1}{T}\int_a^{a+T} U_t f dt.$$

According to the Riesz-Fischer theorem we have to prove that

$$\|\varphi_{T',a'} - \varphi_{T,a}\| \longrightarrow 0,\ T, T' \longrightarrow \infty.$$

We have

$$\|\varphi_{S,b} - \varphi_{T,a}\|^2 = \begin{array}{c}(\varphi_{S,b}, \varphi_{S,b}) + (\varphi_{T,a}, \varphi_{T,a}) \\ -\overline{(\varphi_{S,b}, \varphi_{T,a}) - (\varphi_{S,b}, \varphi_{T,a})},\end{array} \tag{3}$$

where the expressions on the right-hand side may be evaluated as follows.

$$\|\varphi_{S,b} - \varphi_{T,a}\|^2 = \frac{1}{ST}\left(\int_b^{b+S} U_s f ds,\ \int_a^{a+T} U_t f dt\right) = \frac{1}{ST}\int_b^{b+S}\int_a^{a+T} (U_s f, U_t f)\, dt ds.$$

By setting

$$g(\tau) = (U_t f, f),\ F(\tau) = \frac{1}{\tau}\int_0^\tau g(t)dt,\ G(\tau) = \frac{1}{2}\left(F(\tau) + \overline{F(\tau)}\right) \quad (4)$$

and observing that $(U_s f, U_t f) = (U_{s-t} f, f) = g(s-t)$, we readily obtain

$$(\varphi_{S,b}, \varphi_{T,a}) = \frac{1}{ST}\int_{S-T-\delta}^{S-\delta} \tau F(\tau)d\tau - \frac{1}{ST}\int_{-T-\delta}^{-\delta} \tau F(\tau)d\tau;\ \delta = a - b.$$

The second term on the right may be transformed by taking into account the equations

$$g(\tau) = (f, U_{-\tau} f) = \overline{(U_{-\tau} f, f)} = \overline{g(-\tau)},\ f(\tau) = \overline{f(-\tau)}.$$

We obtain

$$(\varphi_{S,b}, \varphi_{T,a}) = \frac{1}{ST}\int_{S-T-\delta}^{S-\delta} \tau F d\tau + \frac{1}{ST}\int_{\delta}^{\delta+T} \tau \bar{F} d\tau;\ \delta = a - b.$$

The application of this equation to the four terms on the right-hand side in (3) leads to our final formula

$$\frac{1}{2}\left\| \varphi_{S,b} - \varphi_{T,a} \right\|^2 = \begin{bmatrix} \frac{1}{S^2}\int_0^S & + \frac{1}{T^2}\int_0^T \\ -\frac{1}{ST}\int_{S-T-\delta}^{S-\delta} & - \frac{1}{ST}\int_{\delta}^{T+\delta} \end{bmatrix} G(\tau)\tau d\tau. \quad (5)$$

We now proceed as follows. According to the Schwarz inequality $g(\tau)$ is a bounded function, $|g| \leqq \|f\| \cdot \|U_\tau f\| = \|f\|^2 = \gamma$,

$$|g(\tau)| \leqq \gamma,\ |G(\tau)| \leqq \gamma. \quad (6)$$

We have

$$(S - \delta - s)G(S - \delta - s) = SG(S) - \int_{S-\delta-s}^{S} g(\tau)\, d\tau$$
$$\geqq SG(S) - \gamma(|s| + |\delta|),$$

thus yielding

$$\int_{S-T-\delta}^{S-\delta} G\tau d\tau = \int_0^T (S - \delta - s)G(S - \delta - s)ds \quad (7)$$
$$\geqq STG(S) - \frac{1}{2}\gamma\,(T + |\delta|)^2.$$

Furthermore we have

$$\left| \int_{\delta}^{T+\delta} G\tau\, d\tau \right| \leqq \frac{1}{2}\gamma\,(T + |\delta|)^2. \quad (8)$$

If we set

$$G^* = \lim_{\tau = \infty} \sup G(\tau)$$

the inequality $G(\tau) < G^* + \frac{1}{8}\epsilon^2$, $\epsilon$ being a given positive number, holds for all sufficiently large values of $\tau$. Hence we conclude that

$$\frac{1}{\tau^2}\int_0^\tau G(\tau)\tau d\tau < \frac{1}{2}\left(G^* + \frac{\epsilon^2}{8}\right), \tau > N(\epsilon). \tag{9}$$

From (5), (7), (8) and (9)

$$\frac{1}{2}\left\|\varphi_{S,b} - \varphi_{T,a}\right\|^2 \leqq G^* - G(S) + \frac{\epsilon^2}{8} + \gamma\frac{(T + |\delta|)^2}{ST} \tag{10}$$

if $S, T < N(\epsilon)$. According to the inequality $\frac{1}{2}\|f_1 + f_2\|^2 \leqq \|f_1\|^2 + \|f_2\|^2$ we have

$$\left\|\varphi_{T',a'} - \varphi_{T,a}\right\|^2 \leqq 2\left\|\varphi_{S,b} - \varphi_{T,a}\right\|^2 + 2\left\|\varphi_{S,b} - \varphi_{T',a'}\right\|^2$$

and according to (10), $\delta = a - b$, $\delta' = a' - b$,

$$\left\|\varphi_{T',a'} - \varphi_{T,a}\right\|^2 \leqq \epsilon^2 + 8(G^* - G(S)) + \frac{4}{S}\gamma\left\{\frac{(T + |\delta|)^2}{T} + \frac{(T' + |\delta'|)^2}{T'}\right\}$$

for $S, T, T' > N(\epsilon)$. If $S \longrightarrow \infty$ in such a way that $G(S) \longrightarrow G^*$, we finally obtain

$$\left\|\varphi_{T',a'} - \varphi_{T,a}\right\| \leqq \epsilon, \text{ if } T, T' > N(\epsilon),$$

which completes the proof.

From (2) we infer the following property of the linear operators $\varphi_{T,a}$,

$$U_\tau \varphi_{T,0} = \varphi_{T,\tau}.$$

This shows that $f^* = U_\tau f^*$ holds for any $\tau$, i.e., that $f^*$ is invariant under the group $U_t$. For an arbitrary invariant point $h$ of the Hilbert space we obtain the equations

$$(\varphi_{T,0}(f), h) = \frac{1}{T}\int_0^T (U_t f, h) = \frac{1}{T}\int_0^T (f, U_{-t}h)dt = (f, h),$$

thus yielding

$$(f^*, h) = (f, h). \tag{11}$$

(11) holds for any invariant $h$ and determines $f^*$ uniquely.

3. *Remarks on the Quasiergodic Dynamical Case.*—In the special case of a dynamical system equation (11) has been used by v. Neumann (loc. cit.) for the determination of the time average $f^*(P)$ of a function $f(P)$ by means of a "space average." The quasiergodic case, which plays a fundamental rôle in classical statistical mechanics, is characterized by the condition that either $m(A) = 0$ or $m(A) = m(\Omega)$ holds for any measurable and invariant subset $A$ of $\Omega$, $m(\Omega)$ being finite. In this case a measurable and invariant function $h(P)$ must be constant almost everywhere on $\Omega$, and (11) yields, $h = 1, f^* = \text{const.}$

$$f^* = \frac{\int_\Omega f\,dm}{m(\Omega)}. \tag{12}$$

A classical case of quasiergodicity is the following. Let $x_1, x_2, \ldots, x_n$ be the cartesian coördinates of a point of the $n$-dimensional Euclidean space. If two points, the corresponding coördinates of which differ by integer multiples of $2\pi$, are identified, we obtain the points of the $n$-dimensional torus, where $x_1, x_2, \ldots, x_n$ now represent the angular coördinates of $P$. We then define $P_t$ by the coördinates

$$x_1 + a_1 t, x_2 + a_2 t, \ldots, x_n + a_n t,$$

where the numbers $a_1, a_2, \ldots, a_n$ are linearly independent. This flow on the torus possesses an invariant measure $m$. From the Fourier series

$$f \sim \sum a_{\nu_1,\nu_2,\ldots\nu_n} \exp i(\nu_1 x_1 + \nu_2 x_2 + \ldots + \nu_n x_n)$$

of a function with finite $\|f\|$ we infer

$$U_t f \sim \sum b_{\nu_1,\nu_2,\ldots\nu_n} \exp.\, i(\nu_1 x_1 + \ldots + \nu_n x_n),$$

$$b_{\nu_1,\nu_2,\ldots\nu_n} = a_{\nu_1\nu_2\ldots\nu_n} \exp.\, it(\nu_1 a_1 + \ldots + \nu_n a_n),$$

which immediately shows that an invariant function must be constant almost everywhere.

The above case has been treated in a well-known paper by Weyl[4] who showed that $\lim_{T=\infty} \frac{1}{T} \int_0^T f(P_t)dt$ exists for all points $P$, if $f(P)$ is continuous everywhere, or if $f$ is the characteristic function of an open point set. This perfect uniformity of the flow, however, is a special feature of the above case. In the general quasiergodic case the exceptional points of Birkhoff's theorem exist.

4. *On Birkhoff's Time Average Theorem.*—Professor Birkhoff's proof of the existence of the time average in the sense of actual convergence is

based upon entirely different considerations of very general nature. In Birkhoff's paper, however, their generality is somewhat obscured by the usage of a manifold of section. The results appear in the following general form, if the proof is arranged in the following way.

Let $P \longrightarrow P_1$ be a single one to one transformation of $\Omega$ into itself which preserves the measure $m$. $m(\Omega)$ is supposed to be finite. We set

$$\phi_n(P) = \frac{1}{n}\,(f(P) + f(P_1) + \ldots + f(P_{n-1})),$$

where $f(P)$ is a bounded and measurable function, and where $P_2 = (P_1)_1$, .... The main point of Birkhoff's method is the

*Lemma.* If

$$\limsup_{n=\infty} \phi_n(P) \geqq \lambda$$

holds everywhere on an invariant and measurable set $A \subset \Omega$, then

$$\int_A f(P)dm \geqq \lambda m(A).$$

The proof is almost exactly the same as in Birkhoff's paper. From the lemma we easily infer the inequality

$$\int_\Omega \limsup_{n=\infty} \phi_n(P)dm \leqq \int_\Omega f(P)dm.$$

In replacing $f$ by $1-f$ we obtain

$$\int_\Omega \liminf_{n=\infty} \phi_n(P)dm \geqq \int_\Omega f(P)dm.$$

This yields immediately the

THEOREM. The limit

$$\lim_{n=\infty} \phi_n(P)$$

exists almost everywhere on $\Omega$.

Let, now, $P \longrightarrow P_t$, $(P_t)_s = P_{t+s}$ be again a group of the above kind and let $f(P)$ be a bounded and measurable function. In setting

$$g(P) = \int_0^1 f(P_t)dt$$

we infer by applying the theorem to $P \longrightarrow P_1$ and to $g(P)$,

$$\phi_n(P) = \frac{1}{n}\int_0^n f(P_t)dt.$$

Taking account to the boundedness of $f(P)$ we obtain immediately Birkhoff's general

TIME AVERAGE THEOREM. For any bounded and measurable function $f(P)$ the limit

$$\lim_{T=\infty} \frac{1}{T} \int_0^T f(P_t)dt$$

exists on $\Omega$ apart from a set of points $P$ of measure zero.

* INTERNATIONAL RESEARCH FELLOW.

[1] Cf. B. O. Koopman, These PROCEEDINGS, **17**, 315–318 (1931).

[2] Cf. J. v. Neumann, These PROCEEDINGS, **18**, 70–82(1932).

[3] Cf. G. D. Birkhoff, These PROCEEDINGS, **17**, 656–660(1931).

[4] Cf. H. Weyl, *Math. Annalen*, **77**, 313–352 (1916).

---

## *SETS OF DISTINCT GROUP OPERATORS INVOLVING ALL THE PRODUCTS BUT NOT ALL THE SQUARES*

BY G. A. MILLER

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS

Communicated November 16, 1931

Suppose that $s_1, s_2, \ldots, s_k$ represent a set of distinct operators which obey the group laws when they are combined and include the product of every pair thereof, irrespective of the order of the factors, but not necessarily the square of any operator of the set. It is well known that a necessary and sufficient condition that such a set constitutes a group is that it contains also the square of each of its elements. In all cases the set generates a group $G$ of order $g \geqq k$. In what follows it will be assumed that $g$ is finite, and we shall determine all the possible groups of finite order which have the property that it is possible to find in each of them a set of distinct operators which generate the group and include the product of every pair of the set, irrespective of the order of the factors, but not the square of all of them. In other words, for each of these sets $k < g$.

When $k = 2$ the set is obviously composed of an arbitrary operator and the identity, and a necessary and sufficient condition is that $G$ is cyclic and this operator is an arbitrary generator of this cyclic group. In this case $k = g$ when $g = 2$ and only then. When $k = 3$ the three operators $s_1, s_2, s_3$ must be commutative and if one of them is the identity, the other two are inverses of each other but are not otherwise restricted. Hence in this case $G$ must again be cyclic, but its order is unrestricted except that it must exceed 2. A necessary and sufficient

condition the $g = k$ in this case is that G is the group of order 3. When none of the three operators $s_1$, $s_2$, $s_3$ is the identity, then each of them must be of order 2 and $G$ is the non-cyclic group of order 4, or the trirectangular group. This is a special case of the following elementary theorem: *If a set of k distinct operators of order 2 has the property that it includes the product of every pair of them, then these operators generate the abelian group of order* $k + 1 = 2^m$ *and of type* (1,1,1,. . . ), *and every such group involves such a set.* Since these cases are very elementary, it may be assumed in what follows that $k > 3$, and that the set $s_1$, $s_2$,. . . ,$s_k$ involves operators whose order exceeds 2.

Suppose that this set involves an operator $s$ and also $s^n \neq s^{-1}$. If $n = 2$ the set involves all the powers of $s$ including the identity. If $n \neq 2$ but less than the order of $s$ diminished by unity the set must involve also $s^{n+1}$, $s^{n+2}$, . . .,$s^{-1}$. Hence it involves also $s^{n-1}$, $s^{n-2}$, . . .,$s^2$. That is, when the set $s_1$, $s_2$, . . .,$s_k$ involves an operator $s$ and also a lower positive power of $s$ than the inverse of $s$ than this set involves also all the powers of $s$ including the identity. In particular, when $G$ is a cyclic group and $k > 3$, then $k = g$. It should be noted that whenever the set $s_1$, $s_2$, . . .,$s_k$ involves an operator, $s$ which is not a power of $s_1$, then it includes also $s_1^n s$ and $s s_1^n$, where $n$ is arbitrary. If it involves $s$ but not $s^2$, each of the operators of the set except possibly the identity must therefore have the property that it generates $s^2$. Hence there results the following theorem: *If a set of distinct operators has the property that it includes the product of every pair of them irrespective of the order of the two factors but does not involve the square of some one of them, then this square must be generated by each of its elements except possibly the identity.*

From this theorem it follows directly that if such a set does not involve the square of some one of its operators, it also cannot involve the square of any other one except possibly the identity, for if the square of such a second operator would then appear in the set, all the powers of this second operator would also appear therein. This is, however, impossible since such powers would involve the square of the first operator. It also results from the theorem noted at the close of the preceding paragraph, that if the square of an element of such a set does not appear therein, then this square must appear in the central of $G$ and hence the square of every element of this set must appear in this central. In particular, when $g > k$ and $G$ is non-abelian than its central quotient group is the abelian group of order $2^m$ and of type (1,1,1, . . . )

It is now easy to prove that when $g > k$ then $G$ cannot be a non-cyclic abelian group unless it is the abelian group of order $2^m$ and of type (1,1,1, . . . ). If $G$ were any other non-cyclic abelian group, the set of operators $s_1$, $s_2$,. . ., $s_k$ would involve at least one operator $s$ such that $s^2$ is not generated by every other operator of the set and hence it would involve all the powers of $s$.

In particular, we could take for such an $s$ one of the independent generators of $G$ of largest order. If an operator of the set which can be used as a second independent generator of $G$ after $s$ has been chosen as a first independent generator does not have its square in the set, then this square can be generated by $s$. As this second operator can clearly be so selected that either its square is in the set or that this square is not generated by $s$, the following theorem has been proved: *If a set of distinct commutative operators contains the product of every pair of them and generates a non-cyclic group, then this set involves all the operators of this group unless each of these operators is of order* 2.

It remains to consider the case when $G$ is non-abelian and $g > k$. It may first be noted that none of the operators of the set $s_1, s_2, \ldots, s_k$ except possibly the identity can be of odd order since it would then be generated by its square and hence also by every operator of this set besides possibly the identity. As it would also generate the square of each of these operators and be commutative with each of them, it results that an operator of order 2 would appear in this set. As this is impossible, it results that besides possibly the identity, each of the operators $s_1, s_2, \ldots, s_k$ is of even order. This order must be the same for all of them and cannot exceed 4. From this there results the following theorem: *If a set of distinct operators has the property that it involves the product of every pair of them irrespective of the order of the two factors, and generates a non-abelian group but does not involve all the operators of this group, then it must be composed of the identity and the six operators of order* 4 *of the quaternion group.*

Any set of $k$ distinct group operators is said to constitute a group whenever it involves the product of every pair of these operators, irrespective of the order of these two factors, and the square of every one of them. From what precedes it results that with the exception of a relatively small number of very elementary groups the latter of these two conditions is implied by the former. This is always the case when some of the operators of the set are non-commutative except when they generate the quaternion group but do not include the operator of order 2 contained in this group. In this case $k = 7$ and $g = 8$. When all the operators of the set are of order 2, they must be commutative, since the product of two operators of order 2 cannot be of order 2 unless these two operators are commutative. This is the only case when $g > k$ and the set of operators $s_1, s_2, \ldots, s_k$ does not include the identity. In all the possible cases when $k < g$ the value of $k$ must be either 2 or of the form $2^m - 1$. When $k = 2$ or 3 the number of distinct groups is infinite; when $m = 3$ there are two possible groups, but for every larger value of $m$ there is one and only one such group.

# CONFORMAL TENSORS

(*First Note*)

BY TRACY YERKES THOMAS

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated December 14, 1931

1. In his treatment of the problem of the equivalence of two quadratic differential forms, E. B. Christoffel[1] was led in 1869 to a sequence of invariants which we now designate as the curvature tensor and its successive covariant derivatives. The process of forming any invariant of this sequence after the first, or curvature tensor, is known as covariant differentiation and can be applied to any arbitrarily given tensor to deduce a tensor of higher rank; however, the significance and indeed the justification of this process of covariant differentiation lies in the fact that it leads to the construction of a sequence of invariants suitable for the complete algebraic characterization of the quadratic differential form, or of the Riemann space, to adopt a more geometrical terminology. We mention this fact in particular because the point of view hereby expressed is fundamental in the following work.

The analogous problem of constructing a sequence of tensor invariants capable of characterizing the *conformal Riemann space* has been recognized for a number of years among geometers. Various aspects of this problem have been treated by Cartan, Schouten, J. M. Thomas, Veblen, Weyl and others. The following discussion will, however, be related primarily to two of my previous notes in These PROCEEDINGS and on this account we shall give a brief statement of the contents of these notes.[2] In the first of these it was recognized that the underlying analytical theory of the conformal Riemann space could be described as the invariant theory of a quadratic differential form

$$\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} G_{\alpha\beta}\,dx^{\alpha}\,dx^{\beta}$$

of weight $-\frac{2}{n}$, where $n$ denotes the dimensionality number of the space; the second note led to the definition of a connection with components ${}^{0}\Gamma^{i}_{j\alpha}$ obeying a law of transformation

$$ {}^{0}\bar{\Gamma}^{q}_{j\alpha}\,u^{i}_{q} = \frac{\partial u^{i}_{j}}{\partial x^{\alpha}} + {}^{0}\Gamma^{i}_{p\mu}\,u^{p}_{j}\,u^{\mu}_{\alpha} \qquad (1.1)$$

under an arbitrary analytic transformation of the $x$ coördinates. In the above equations (1.1), as in all following equations, we adopt the sum-

mation convention with the understanding that Greek indices assume the values 0, 1, ..., $n$ and Latin indices the values 0, 1, ..., $n$, $\infty$ unless the contrary is indicated; here $\infty$ is introduced as a convenient symbol in place of $n + 1$ as used by some writers. The quantities $u^i_k$ in (1.1) are defined as the elements of the matrix

| | 0 | 1 | . . . | $n$ | $\infty$ |
|---|---|---|---|---|---|
| 0 | 1 | 0 | . . . | 0 | 0 |
| 1 | $\bar{\psi}_1$ | $\frac{\partial x^1}{\partial \bar{x}^1}$ | . . . | $\frac{\partial x^n}{\partial \bar{x}^1}$ | 0 |
| . | . | . | | . | . |
| . | . | . | | . | . |
| . | . | . | | . | . |
| $n$ | $\bar{\psi}_n$ | $\frac{\partial x^1}{\partial \bar{x}^n}$ | . . . | $\frac{\partial x^n}{\partial \bar{x}^n}$ | 0 |
| $\infty$ | $\frac{1}{2}\bar{\psi}_\sigma\bar{\psi}_\sigma$ | $\frac{\partial x^1}{\partial \bar{x}^\sigma}\bar{\psi}^\sigma$ | . . . | $\frac{\partial x^n}{\partial \bar{x}^\sigma}\bar{\psi}^\sigma$ | $(x\bar{x})^{\frac{2}{n}}$ |

in such a way that $u^i_k$ stands for the element in the i$^{th}$ column and k$^{th}$ row of this matrix; also

$$\bar{\psi}_\sigma = \frac{\partial \log (x\bar{x})}{\partial \bar{x}^\sigma}, \quad \bar{\psi}^\sigma = \bar{G}^{\sigma\tau}\bar{\psi}_\tau$$

$$(\sigma,\tau = 1, \ldots, n)$$

where $(x\bar{x})$ denotes the jacobian determinant of the coördinate transformation. As a knowledge of certain of the components ${}^0\Gamma^i_{j\alpha}$ is necessary in our later work we shall observe that these are defined by

$${}^0\Gamma^i_{k0} = -\frac{1}{n}\delta^i_k; \quad {}^0\Gamma^i_{0\gamma} = -\frac{1}{n}\delta^i_\gamma,$$

$${}^0\Gamma^\alpha_{\beta\gamma} = K^\alpha_{\beta\gamma}; \quad {}^0\Gamma^0_{\beta\gamma} = \left(\frac{n}{n-2}\right)Q_{\beta\gamma}; \quad {}^0\Gamma^\infty_{\beta\gamma} = -\frac{1}{n}G_{\beta\gamma}, \quad (\alpha,\beta,\gamma = 1,\ldots,n)$$

$${}^0\Gamma^\alpha_{\infty\beta} = \left(\frac{n}{n-2}\right)Q^\alpha_\beta; \quad {}^0\Gamma^0_{\infty\gamma} = {}^0\Gamma^\infty_{\infty\gamma} = 0. \quad (\alpha,\beta,\gamma = 1,\ldots,n).$$

Reference may be had to the second of the above-mentioned papers for the explicit definition of the quantities $Q$ and $K$ in the above expressions. The quantities $G_{\alpha\beta}(\alpha,\beta = 1, \ldots, n)$ which have the law of transformation

$$\bar{G}_{\alpha\beta} = (x\bar{x})^{-\frac{2}{n}} G_{\mu\nu}\, u^\mu_\alpha\, u^\nu_\beta \quad \text{(indices} = 1, \ldots, n) \qquad (1.2)$$

constitute the components of the *fundamental conformal tensor*. In addition to the above equations we shall also need the equations

$$\overline{G}^{\alpha\beta} = (x\bar{x})^{\frac{2}{n}} G^{\mu\nu} u^{\alpha}_{\mu} u^{\beta}_{\nu} \quad (\text{indices} = 1, \ldots, n) \qquad (1.3)$$

which give the law of transformation of the contravariant components $G^{\mu\nu}$ of the fundamental conformal tensor.

It is evident that equations (1.1) can be used in the usual manner to construct from an arbitrarily given tensor $T$, a covariant derivative for which the values 0, 1, ..., $n$ only will be assumed by the added index of the differentiation; on account of the restricted range of this index, this process of covariant differentiation cannot, however, be used to form a second covariant derivative of tensor character. This suggests that we "complete" the above covariant derivative by defining the values of its components for the value $\infty$ of the index of differentiation.[3] Using the idea of completing this covariant derivative we introduce in this note a method which succeeds for all cases for which a certain constant $K$ does not vanish, the constant $K$ depending on the dimensionality number $n$ of the space, the number of contravariant indices, the number of covariant indices and the weight of the given tensor. An extension of this method which enables us to complete a tensor whose components involve two Greek indices of restricted range 0, 1, ..., $n$ is given in § 3. Corresponding to the above case, this latter method fails to apply whenever another constant $L$, depending on the same quantities as $K$, is equal to zero. This method can in general be applied to the conformal curvature tensor as defined in the second of my above-mentioned notes in These PROCEEDINGS, since for this tensor $L = 0$ only if $n = 4$; hence we arrive at covariant derivatives of the conformal curvature tensor having the significance that the equations of transformation of their components express integrability conditions of (1.1). It turns out that if the dimensionality number $n$ is odd our method leads to the formation of an infinite sequence of covariant derivatives of the conformal curvature tensor. For $n$ even but different from 4, the constant $K$ will always vanish at some stage of the process of forming this sequence so that our method of constructing successive covariant derivatives cannot be continued.

A slight formal modification of a well-known lemma on partial differential equations[4] suffices to show that, for $n$ odd, the conformal curvature tensor and its successive covariant derivatives constitute, when combined with the fundamental conformal tensor $G$, a complete set of tensor invariants of the conformal Riemann space.

In a second note on this same subject we shall treat the exceptional cases $K = 0$ and $L = 0$ on the basis of the fundamental idea of covariant differentiation, namely, the construction of conditions of integrability

in tensorial form of the equations of transformation of the components of a tensor.

2. *Covariant derivative of an arbitrary tensor T.* DEFINITION. *A complete relative conformal tensor T of weight W or simply a relative conformal tensor of weight W, is an entity with components* $T^{p\cdots q}_{r\cdots s}$ *which depend on the coördinates* $x^1, \ldots, x^n$ *alone and which have the transformation*

$$T^{i\cdots j}_{k\cdots l}\, u^p_i \ldots u^q_j = \left| u^a_b \right|^W T^{p\cdots q}_{r\cdots s}\, u^r_k \ldots u^s_l \qquad (2.1)$$

*when the coördinates undergo an arbitrary analytic transformation whose jacobian determinant does not vanish identically. If, however, one or more of the indices in these equations are limited to the restricted range* 0, 1, ..., *n the entity T so defined is called an incomplete conformal tensor of weight W.* The quantity $\left| u^a_b \right|$ in (2.1) denotes the determinant of order $n + 2$ formed from the quantities $u^a_b$ and in fact is such that

$$\left| u^a_b \right| = (\bar{x}x)^{\frac{n+2}{n}}.$$

Observing first that

$$\frac{\partial \left| u^a_b \right|}{\partial \bar{x}^\alpha} = W \left| u^a_b \right|^W \left\{ {}^0\Gamma^h_{h\alpha} - {}^0\Gamma^h_{h\sigma}\, u^\sigma_\alpha \right\},$$

we find by differentiating (2.1) with respect to $\bar{x}^\alpha$ and eliminating derivatives of the $u^i_k$ which occur, by means of (1.1), that

$$\bar{T}^{i\cdots j}_{k\cdots l,\alpha}\, u^p_i \ldots u^q_j = \left| u^a_b \right|^W T^{p\cdots q}_{r\cdots s,\mu}\, u^r_k \ldots u^s_l\, u^\mu_\alpha, \qquad (2.2)$$

where

$$T^{p\cdots q}_{r\cdots s,\mu} = \frac{\partial T^{p\cdots q}_{r\cdots s}}{\partial x^\mu} + T^{h\cdots q}_{r\cdots s}\, {}^0\Gamma^p_{h\mu} + \ldots + T^{p\cdots h}_{r\cdots s}\, {}^0\Gamma^q_{h\mu} - T^{p\cdots q}_{h\cdots s}\, {}^0\Gamma^h_{r\mu} - \ldots - T^{p\cdots q}_{r\cdots h}\, {}^0\Gamma^h_{s\mu} - WT^{p\cdots q}_{r\cdots s}\, {}^0\Gamma^h_{h\mu}, \qquad (2.3)$$

and there is an analogous expression for the components appearing in the left members of (2.2); in the following it will be assumed without special mention that whenever a set of quantities such as $T^{p\cdots q}_{r\cdots s,\mu}$ are defined by equations of the type (2.3) the corresponding barred quantities are defined in an analogous manner.

Let $N$ denote the number of indices $p \ldots q$ and $M$ the number of indices $r \ldots s$ appearing in the components of the above tensor $T$. Then (2.3) show that

$$T^{p\cdots q}_{r\cdots s,0} = \left[ \frac{M - N + (n+2)W}{n} \right] T^{p\cdots q}_{r\cdots s}. \qquad (2.4)$$

Now we can form a conformal tensor from the incomplete conformal tensor whose components occur in (2.2) if we can define a set of quantities

$T^{p\cdots q}_{r\cdots s,\omega}$ depending on the coördinates $x^1, \ldots, x^n$ alone and having equations of transformation

$$\overline{T}^{i\cdots j}_{k\cdots l,\omega}\, u^p_i \ldots u^q_j = |u^a_b|^W \Big\{ T^{p\cdots q}_{r\cdots s,\omega}\, u^\omega_\omega + T^{p\cdots q}_{r\cdots s,0}\, u^0_\omega + \sum_{\sigma=1}^{n} T^{p\cdots q}_{r\cdots s,\sigma}\, u^\sigma_\omega \Big\}\ u^r_k \ldots u^s_l. \quad (2.5)$$

In fact (2.2) and (2.5) combine to give

$$\overline{T}^{i\cdots j}_{k\cdots l,f}\, u^p_i \ldots u^q_j = |u^a_b|^W\, T^{p\cdots q}_{r\cdots s,g}\, u^r_k \ldots u^s_l\, u^g_f. \quad (2.6)$$

The conformal tensor whose components appear in (2.2) will be called the incomplete covariant derivative and the tensor whose components appear in (2.6), the complete covariant derivative, or simply the covariant derivative of the given tensor $T$.

To construct the desired quantities $T^{p\cdots q}_{r\cdots s,\omega}$ transforming by (2.5) we first differentiate (2.2) with respect to $\bar{x}^\beta$, obtaining

$$\overline{T}^{i\cdots j}_{k\cdots l,\alpha\beta}\, u^p_i \ldots u^q_j = |u^a_b|^W \Big\{ T^{p\cdots q}_{r\cdots s,\mu\nu}\, u^\mu_\alpha\, u^\nu_\beta + T^{p\cdots q}_{r\cdots s,\sigma}\, {}^0\overline{\Gamma}^\omega_{\alpha\beta}\, u^\sigma_\omega \Big\}\, u^r_k \ldots u^s_l, \quad (2.7)$$

where

$$\begin{aligned} T^{p\cdots q}_{r\cdots s,\mu\nu} = \frac{\partial T^{p\cdots q}_{r\cdots s,\mu}}{\partial x^\nu} &+ T^{h\cdots q}_{r\cdots s,\mu}\, {}^0\Gamma^p_{h\nu} + \ldots + T^{p\cdots h}_{r\cdots s,\mu}\, {}^0\Gamma^q_{h\nu} \\ &- T^{p\cdots q}_{h\cdots s,\mu}\, {}^0\Gamma^h_{r\nu} - \ldots - T^{p\cdots q}_{r\cdots h,\mu}\, {}^0\Gamma^h_{s\nu} \\ &- T^{p\cdots q}_{r\cdots s,\sigma}\, {}^0\Gamma^\sigma_{\mu\nu} - W T^{p\cdots q}_{r\cdots s,\mu}\, {}^0\Gamma^h_{h\nu}. \end{aligned}$$

In particular we can deduce from these latter equations and (2.4) that

$$T^{p\cdots q}_{r\cdots s,0\nu} = \left[\frac{M - N + (n+2)W + 1}{n}\right] T^{p\cdots q}_{r\cdots s,\nu}, \quad (2.8)$$

$$T^{p\cdots q}_{r\cdots s,\mu 0} = \left[\frac{M - N + (n+2)W + 1}{n}\right] T^{p\cdots q}_{r\cdots s,\mu}, \quad (2.9)$$

$$T^{p\cdots q}_{r\cdots s,00} = \left[\frac{M - N + (n+2)W + 1}{n}\right]\left[\frac{M - N + (n+2)W}{n}\right] T^{p\cdots q}_{r\cdots s}. \quad (2.10)$$

Now restrict the values of $\alpha$ and $\beta$ in (2.7) to the range $1, \ldots, n$ and rewrite the right members of these equations so as to obtain a corresponding restriction of range for the indices $\mu$ and $\nu$, i.e., we form the equations

$$\begin{aligned} \overline{T}^{i\cdots j}_{k\cdots l,\alpha\beta}\, u^p_i \ldots u^q_j = |u^a_b|^W \Big\{ & T^{p\cdots q}_{r\cdots s,\mu\nu}\, u^\mu_\alpha\, u^\nu_\beta + T^{p\cdots q}_{r\cdots s,0\nu}\, u^0_\alpha\, u^\nu_\beta \\ & + T^{p\cdots q}_{r\cdots s,\mu 0}\, u^\mu_\alpha u^0_\beta + T^{p\cdots q}_{r\cdots s,00}\, u^0_\alpha\, u^0_\beta \\ & + T^{p\cdots q}_{r\cdots s,\nu}\, {}^0\overline{\Gamma}^\omega_{\alpha\beta}\, u^\nu_\omega + T^{p\cdots q}_{r\cdots s,0}\, {}^0\overline{\Gamma}^\omega_{\alpha\beta}\, u^0_\omega \Big\}\, u^r_k \ldots u^s_l \quad (2.11) \end{aligned}$$

with the understanding that the indices $\alpha,\beta,\mu,\nu$ assume values 1, ..., $n$. If we multiply (2.11) by $\overline{G}^{\alpha\beta}$ and sum on the indices $\alpha$ and $\beta$, taking account of (1.3), (2.4), (2.8), (2.9) and (2.10) we obtain a system of equations which can be written

$$\overline{T}^{i\ldots j}_{k\ldots l,\alpha\beta}\,\overline{G}^{\alpha\beta}\,u^p_i \ldots u^q_j = |u^a_b|^W \left\{ T^{p\ldots q}_{r\ldots s,\mu\nu}\,G^{\mu\nu}\,u^{\infty}_{\infty} + \left[\frac{2M-2N+2(n+2)W+2-n}{n}\right](T^{p\ldots q}_{r\ldots s,0}\,u^0_{\infty} + T^{p\ldots q}_{r\ldots s,\nu}\,u^{\nu}_{\infty})\right\} u^r_k \ldots u^s_l. \tag{2.12}$$

Hence, if the constant $K$ defined by

$$K = 2M - 2N + 2(n+2)W + 2 - n$$

does not vanish, we can put

$$T^{p\ldots q}_{r\ldots s,\infty} = \left(\frac{n}{K}\right)\sum_{\mu=1}^{n}\sum_{\nu=1}^{n} T^{p\ldots q}_{r\ldots s,\mu\nu}\,G^{\mu\nu}, \tag{2.13}$$

and the components $T^{p\ldots q}_{r\ldots s,\infty}$ so defined will transform by (2.5). We thus obtain the complete covariant derivative with components $T^{p\ldots q}_{r\ldots s,\xi}$ transforming in accordance with (2.6).

If $K = 0$, equations (2.12) show the existence of a conformal tensor $\mathfrak{T}$ of weight $W + \dfrac{2}{n+2}$ having the components

$$\sum_{\mu=1}^{n}\sum_{\nu=1}^{n} T^{p\ldots q}_{r\ldots s,\mu\nu}\,G^{\mu\nu}.$$

The constant $K$ will not vanish for the tensor $\mathfrak{T}$ so that it is possible to construct the complete conformal covariant derivative of this tensor. It is, in fact, evident that we can form the infinite sequence of successive covariant derivatives of the tensor $\mathfrak{T}$.

3. *An Extension of the Preceding Method.*—Before applying the above theory to the problem of constructing conformal tensor invariants suitable for the characterization of the conformal space we must consider the problem of completing a tensor $D$ having components $D^{p\ldots q}_{r\ldots s\mu\nu}$ which are skew-symmetric in the indices $\mu$ and $\nu$ and such that

$$D^{p\ldots q}_{r\ldots s0\nu} \equiv D^{p\ldots q}_{r\ldots s\mu 0} \equiv 0.$$

Differentiation of the equations of transformation of the components of the tensor $D$, namely,

$$\overline{D}^{i\ldots j}_{k\ldots l\alpha\beta}\,u^p_i \ldots u^q_j = |u^a_b|^W\, D^{p\ldots q}_{r\ldots s\mu\nu}\,u^r_k \ldots u^s_l\,u^u_\alpha\,u^v_\beta$$

and elimination of the derivatives of the $u^i_k$ which occur, by (1.1), gives

$$\overline{D}^{i\cdots j}_{k\cdots l\alpha\beta\gamma}\, u^p_i \cdots u^q_j = |u^a_b|^W \left\{ D^{p\cdots q}_{r\cdots s\mu\nu\pi}\, u^\mu_\alpha u^\nu_\beta u^\pi_\gamma + D^{p\cdots q}_{r\cdots s\mu\nu} ({}^0\overline{\Gamma}^\infty_{\alpha\gamma}\, u^\mu_\infty\, u^\nu_\beta + {}^0\overline{\Gamma}^\infty_{\beta\gamma}\, u^\mu_\alpha\, u^\nu_\infty) \right\} u^r_k \cdots u^s_l, \tag{3.2}$$

where

$$D^{p\cdots q}_{r\cdots s\mu\nu\pi} = \frac{\partial D^{p\cdots q}_{r\cdots s\mu\nu}}{\partial x^\pi} + D^{h\cdots q}_{r\cdots s\mu\nu}\, {}^0\Gamma^p_{h\pi} + \cdots + D^{p\cdots h}_{r\cdots s\mu\nu}\, {}^0\Gamma^q_{h\pi} - D^{p\cdots q}_{h\cdots s\mu\nu}\, {}^0\Gamma^h_{r\pi} - \cdots - D^{p\cdots q}_{r\cdots h\mu\nu}\, {}^0\Gamma^h_{s\pi} - D^{p\cdots q}_{r\cdots s\sigma\nu}\, {}^0\Gamma^\sigma_{\mu\pi} - D^{p\cdots q}_{r\cdots s\mu\sigma}\, {}^0\Gamma^\sigma_{\nu\pi} - WD^{p\cdots q}_{r\cdots s\mu\nu}\, {}^0\Gamma^h_{h\pi}.$$

If as before $N$ and $M$ denote the number of indices $p\ldots q$ and $r\ldots s$, respectively, then it follows from the last formula that

$$D^{p\cdots q}_{r\cdots s0\nu\pi} = \frac{1}{n} D^{p\cdots q}_{r\cdots s\pi\nu},$$

$$D^{p\cdots q}_{r\cdots s\mu0\pi} = \frac{1}{n} D^{p\cdots q}_{r\cdots s\mu\pi},$$

$$D^{p\cdots q}_{r\cdots s\mu\nu0} = \left[\frac{M - N + (n+2)W + 2}{n}\right] D^{p\cdots q}_{r\cdots s\mu\nu}.$$

Equations (3.2) can accordingly be written

$$\overline{D}^{i\cdots j}_{k\cdots l\alpha\beta\gamma}\, u^p_i \cdots u^q_j = |u^a_b|^W \left\{ D^{p\cdots q}_{r\cdots s\mu\nu\pi}\, u^\mu_\alpha u^\nu_\beta u^\pi_\gamma + \frac{1}{n} D^{p\cdots q}_{r\cdots s\pi\nu}\, u^0_\alpha u^\nu_\beta u^\pi_\gamma + \frac{1}{n} D^{p\cdots q}_{r\cdots s\mu\pi}\, u^\mu_\alpha u^0_\beta u^\pi_\gamma + \left[\frac{M - N + (n+2)W + 2}{n}\right] D^{p\cdots q}_{r\cdots s\mu\nu}\, u^\mu_\alpha u^\nu_\beta u^0_\gamma + D^{p\cdots q}_{r\cdots s\mu\nu} ({}^0\overline{\Gamma}^\infty_{\alpha\gamma}\, u^\mu_\infty\, u^\nu_\beta + {}^0\overline{\Gamma}^\infty_{\beta\gamma}\, u^\mu_\alpha\, u^\nu_\infty) \right\} u^r_k \cdots u^s_l \tag{3.3}$$

in which it is to be understood that the indices $\alpha,\beta,\gamma,\mu,\nu,\pi$ admit only the restricted range $1, \ldots, n$. Multiplying (3.3) by $\overline{G}^{\beta\gamma}$ and summing on the indices $\beta$ and $\gamma$, we obtain after some reductions

$$\overline{D}^{i\cdots j}_{k\cdots l\alpha\beta\gamma}\, \overline{G}^{\beta\gamma}\, u^p_i \cdots u^q_j = |u^a_b|^w \left\{ D^{p\cdots q}_{r\cdots s\mu\nu\pi}\, G^{\nu\pi}\, u^\mu_\alpha\, u^\infty_\infty + \left(\frac{L}{n}\right) D^{p\cdots q}_{r\cdots s\mu\nu}\, u^\mu_\alpha\, u^\nu_\infty \right\} u^r_k \cdots u^s_l \tag{3.4}$$

where the indices $\alpha,\beta,\gamma,\mu,\nu,\pi$ admit the range $1, \ldots, n$ only and the integer $L$ is defined by

$$L \equiv M - N + (n+2)W + 4 - n.$$

Since

$$D^{p\cdots q}_{r\cdots s0\nu\pi}\, G^{\nu\pi} \equiv 0$$

owing to the skew-symmetric property of the components of the incomplete conformal tensor $D$, it follows that (3.4) holds also when regarded as equations for the transformation of the quantities

$$D^{p\cdots q}_{r\cdots s\mu\nu\pi}\ G^{\nu\pi}, \tag{3.5}$$

where the index $\mu$ takes on the full range of values $0, 1, \ldots, n$ in conformity with the general convention; in the following we shall regard (3.4) as holding in this sense.

If $L \neq 0$, we put

$$D^{p\cdots q}_{r\cdots s\mu\infty} \equiv -\ D^{p\cdots q}_{r\cdots s\infty\mu} \equiv \left(\frac{n}{L}\right) D^{p\cdots q}_{r\cdots s\mu\nu\pi}\ G^{\nu\pi},$$

$$D^{p\cdots q}_{r\cdots s0\infty} \equiv D^{p\cdots q}_{r\cdots s\infty 0} \equiv D^{p\cdots q}_{r\cdots s\infty\infty} \equiv 0.$$

Then (3.1) and (3.4) combine to give

$$D^{i\cdots j}_{k\cdots lcd}\ u^p_i \ldots u^q_j = |\ u^a_b\ |^W\ D^{p\cdots q}_{r\cdots sfg}\ u^r_k \ldots u^i_l u^f_c u^g_d,$$

and we have succeeded in completing the incomplete conformal tensor $D$; from their definition the components of the complete conformal tensor $D$ are skew-symmetric in their last two indices, i.e.,

$$D^{p\cdots q}_{r\cdots sfg} \equiv -\ D^{p\cdots q}_{r\cdots sgf}.$$

When $L = 0$, equations (3.4) lead to the definition of an incomplete conformal tensor of weight $W + \dfrac{2}{n+2}$ with components (3.5); it is obvious that a method analogous to the above can in general be used to complete this tensor after which its successive covariant derivatives can be constructed.

4. *An Application of the Method of* § 3.—As a first application of the method of the preceding paragraph we can consider the components of an incomplete conformal tensor $D$ to be defined by

$$D^{p\cdots q}_{r\cdots s\mu\nu} \equiv T^{p\cdots q}_{r\cdots s,\mu\nu} - T^{p\cdots q}_{r\cdots s,\nu\mu}, \tag{4.1}$$

Cf. equation (2.7). The tensor $D$ is then the direct analogue of the skew-symmetric part of the second covariant derivative of a given conformal tensor $T$. It is therefore always possible to complete the tensor $D$ defined by (4.1) even when our method for constructing the complete covariant derivative of the tensor $T$ fails to apply, since the integers $K$ and $L$ cannot vanish simultaneously.

5. *The Complete Conformal Curvature Tensor* ($n \neq 4$).—The important application of the method of § 3 and the one in fact for which this method was designed, is to complete the curvature tensor whose components

${}^0B^i_{k\alpha\beta}$ were deduced in the second of my above mentioned notes in These PROCEEDINGS (see § 1). That is, we put

$$D^i_{k\alpha\beta} \equiv {}^0B^i_{k\alpha\beta};$$

the requisite conditions on the components $D^i_{k\alpha\beta}$ are then satisfied since the components ${}^0B^i_{k\alpha\beta}$ in addition to being skew-symmetric in the indices $\alpha$ and $\beta$, are equal to zero whenever the values $\alpha = 0$ or $\beta = 0$ are assumed. For this case $L = 4 - n$ and hence we can use the method of § 3 to complete the curvature tensor whenever $n$ is different from 4. We thus obtain the components ${}^0B^i_{klm}$ of the complete conformal curvature tensor.

6. *Tensors Derived from the Conformal Curvature Tensor.*—Starting with the complete conformal curvature tensor, $n \neq 4$, let us consider the problem of forming its successive covariant derivatives. We have $N = 1$ and $W = 0$ so that the integer $K$ is equal to $2M - n$ for this case. Hence if $n$ is an odd integer, greater than or equal to 3, we can construct the infinite sequence of complete covariant derivatives of the complete conformal curvature tensor. Let us indicate this by writing

$$\boxed{n \text{ odd}} \quad {}^0B^i_{klm};\ {}^0B^i_{klmp};\ {}^0B^i_{klmpq};\ \ldots.$$

as the components of the curvature tensor and its successive covariant derivatives.

If $n$ is even and greater than 4 the constant $K$ will always vanish at some stage of the process of forming the infinite sequence of complete covariant derivatives of the curvature tensor. For example, if $n = 6$, $K = 0$ for the conformal curvature tensor; if $n = 8$, we can construct the first complete conformal covariant derivative of the conformal curvature tensor but then find that $K = 0$ for this covariant derivative, etc. The following indications for $n = 6$, 8 and 10 enable us to see at a glance the behavior of these sequences for the above and higher values of the dimensionality number.

$$\boxed{n = 6} \quad {}^0B^i_{klm}.$$

$$\boxed{n = 8} \quad {}^0B^i_{klm};\ {}^0B^i_{klmp}.$$

$$\boxed{n = 10} \quad {}^0B^i_{klm};\ {}^0B^i_{klmp};\ {}^0B^i_{klmpq}.$$

As shown these sequences end after a finite number of terms after which our method for completing the covariant derivative fails to apply. Use of the quantities $P_{ik}$ defined in the second of my above-mentioned notes in These PROCEEDINGS will evidently lead to an entirely covariant form of the components of the conformal curvature tensor and its covariant derivatives.

[1] E. B. Christoffel, *J. reine angew. Math.* (*Crelle*), **70**, 46 and 241 (1869).

[2] T. Y. Thomas, "Invariants of Relative Quadratic Differential Forms," These PROCEEDINGS, **11**, 722–725 (1925); also "On Conformal Geometry," *Ibid.*, **12**, 352–359 (1926).

[3] O. Veblen completed the covariant derivative by imposing a set of invariant conditions involving components of conformal tensors. These conditions correspond to our equations (2.13) but differ from these latter equations in the important respect that they involve only derivatives of the first order. See "Differential Invariants and Geometry," *Atti del congresso internazionale dei matematici*, **1**, [6] 181–189 (1928); also, "Conformal Tensors and Connections," These PROCEEDINGS, **14**, 735–745 (1928). Under certain conditions this method enables one to complete the covariant derivative of a tensor $T$ in a very satisfactory manner. However, Veblen's method does not lead to a sequence of conformal tensor invariants, analogous to the Riemann curvature tensor and its successive covariant derivatives, by means of which the equivalence or non-equivalence of two conformal spaces can be determined. On this account the method of the present note is preferable in my opinion.

[4] See, for example, L. P. Eisenhart, *Non-Riemannian Geometry* (New York) 1927, pp. 14–18, where references to the literature are to be found.

---

## *ON THE CHANGES OF SIGN OF THE DERIVATIVES OF A FUNCTION DEFINED BY A LAPLACE INTEGRAL*

BY D. V. WIDDER

DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY

Communicated November 21, 1931

Let a real function $f(x)$ be defined by a Laplace integral

$$f(x) = \int_0^\infty e^{-xt}\,\varphi(t)dt,$$

where $\varphi(t)$ is a real continuous function in the interval $0 \leqq t < \infty$, the integral converging for $x > 0$. The purpose of this note is to announce certain results concerning the changes of sign of $f(x)$ and its derivatives as affected by the changes of sign of $\varphi(t)$.

It was known to Laguerre[1] that $f(x)$ cannot have more changes of sign than $\varphi(t)$. Since

$$f^{(k)}(x) = (-1)^k \int_0^\infty e^{-xt}\,t^k\varphi(t)\,dt,$$

it follows that $f^{(k)}(x)$ cannot have more changes of sign than $\varphi(t)$. We are able to show that it has exactly as many as $\varphi(t)$ for all $k$ sufficiently large. Moreover, we show that if a change of sign of $\varphi(t)$ is at $t = a$, then one of the changes of sign of $f^{(k)}(x)$ will be at a point $x_k$ such that

$$\lim_{k=\infty} \frac{x_k}{k} = \frac{1}{a}.$$

In particular, if $\varphi(t)$ has infinitely many changes of sign, the number experienced by $f^{(k)}(x)$ will become infinite as $k$ becomes infinite. We state the theorems leading up to these results. The proofs of the theorems will be published elsewhere.

THEOREM 1. The function

$$(1 + x)^{-k} - e^{-kx}$$

tends uniformly to zero in the interval $0 \leqq x < \infty$ as $k$ becomes infinite.

COROLLARY. The function

$$\left[\left(\frac{x}{x+n}\right)^{k+1} - e^{-\frac{kn}{x}}\right] \qquad (n = 0, 1, 2, \ldots)$$

tends uniformly to zero in the interval $0 \leqq x < \infty$ as $k$ becomes infinite.

THEOREM 2. If $G_N(x)$ is an exponential polynomial,

$$G_N(x) = \sum_{n=0}^{N} a_n e^{-nx},$$

and

$$f(x) = \int_0^\infty e^{-xt} G_N(t)\, dt,$$

then

$$\lim_{k=\infty}\left[\frac{x^{k+1}(-1)^k f^{(k)}(x)}{k!} - G_N\left(\frac{k}{x}\right)\right]$$

uniformly in the interval $0 \leqq x < \infty$.

In this theorem and in the preceding corollary the function in brackets is to be defined as zero when $x = 0$.

THEOREM 3. If $\varphi(t)$ is continuous in the interval $0 \leqq t < \infty$, tending to a finite limit as $t$ becomes infinite, then to an arbitrary positive $\epsilon$ there corresponds an exponential polynomial

$$G_N(t) = \sum_{n=0}^{N} a_n e^{-nt}$$

such that

$$|\varphi(t) - G_N(t)| < \epsilon \qquad (0 \leqq t < \infty).$$

THEOREM 4. If $\varphi(t)$ satisfies the conditions of Theorem 3, and if

$$f(x) = \int_0^\infty e^{-xt} \varphi(t)\, dt,$$

then $$\lim_{k=\infty}\left[\frac{x^{k+1}(-1)^k f^{(k)}(x)}{k!} - \varphi\left(\frac{k}{x}\right)\right] = 0$$

uniformly in the interval $0 \leqq x < \infty$.

THEOREM 5. If $\varphi(t)$ satisfies the conditions of Theorem 3, if $\varphi(t)$ has $n$ changes of sign in the interval $0 < t < \infty$, and if

$$f(x) = \int_0^\infty e^{-xt}\,\varphi(t)\,dt,$$

then $f^{(k)}(x)$ has exactly $n$ changes of sign in $0 < x < \infty$ for $k$ sufficiently large.

COROLLARY. If $\varphi(t)$ has infinitely many changes of sign, then $f^{(k)}(x)$ has $n_k$ changes of sign in $0 < x < \infty$, where

$$\lim_{k=\infty} n_k = \infty.$$

THEOREM 6. If $\varphi(t)$ satisfies the conditions of Theorem 3, has changes of sign at the points $0 < t_1 < t_2 < \ldots < t_n$ and at no others, is not identically zero in a neighborhood of any of these points $t_i$, and if

$$f(x) = \int_0^\infty e^{-xt}\,\varphi(t)\,dt,$$

then $f^{(k)}(x)$ has exactly $n$ changes of sign for $k$ sufficiently large at the points

$$x_{1k} > x_{2k} > \ldots > x_{nk},$$

and

$$\lim_{k=\infty} \frac{x_{ik}}{k} = \frac{1}{t_i} \qquad (i = 1, 2, \ldots, n).$$

We note that $\varphi(t)$ satisfies the conditions of this theorem if, in particular, it is analytic in the interval $0 \leqq t < \infty$ and approaches a finite limit as $t$ becomes infinite.

The results of Theorems 5 and 6 appear to hold if the restriction that $\varphi(t)$ tends to a finite limit when $t$ becomes infinite is omitted. To establish the theorems in this case would involve a change of the present method since the condition cannot be omitted in Theorem 3.

[1] Laguerre, *Oeuvres*, **1**, p. 28. See also G. Pólya, "Sur un théorème de Laguerre," *Compt. Rend. Séances l'Académie Sci.*, **156**, 996 (1913).

# THE RELATIONSHIP BETWEEN CHEMICAL CONSTITUTION AND TASTE

BY ARTHUR L. FOX

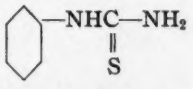
JACKSON LABORATORY, E. I. DU PONT DE NEMOURS & CO., WILMINGTON, DELAWARE

Read before the Academy Tuesday, November 17, 1931

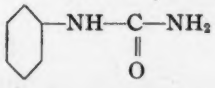
Some time ago the author had occasion to prepare a quantity of phenyl thio carbamide, and while placing it in a bottle the dust flew around in the air. Another occupant of the laboratory, Dr. C. R. Noller, complained of the bitter taste of the dust, but the author, who was much closer, observed no taste and so stated. He even tasted some of the crystals and assured Dr. Noller they were tasteless but Dr. Noller was equally certain it was the dust he tasted. He tried some of the crystals and found them extremely bitter. With these two diverse observations as a starting point, a large number of people were investigated and it was established that this peculiarity was not connected with age, race or sex. Men, women, elderly persons, children, negroes, Chinese, Germans and Italians were all shown to have in their ranks both tasters and non-tasters.

At first it was thought that this phenomenon was connected with the acidity or alkalinity of the mouth, but experiment soon showed there was no connection whatever.

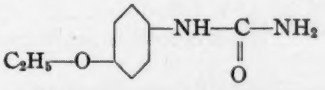
This peculiar phenomenon offered an interesting opportunity for a chemical study of related compounds. Phenyl thio carbamide has the structure



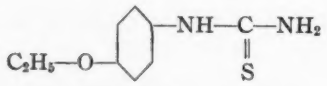
which is closely related to phenyl carbamide, with the structure



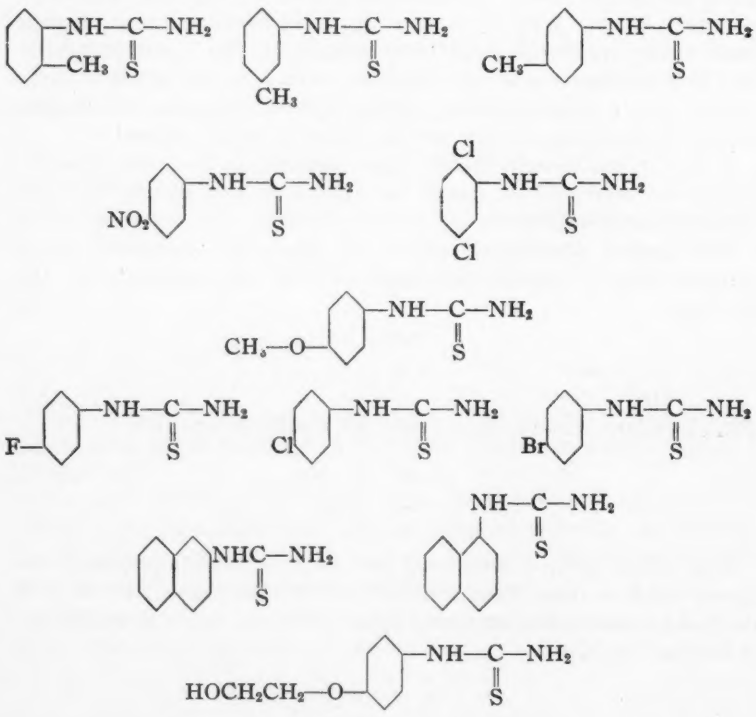
If an ethoxy group is introduced para to the nitrogen a product is obtained which is three hundred times sweeter than sugar, and which is used as a commercial sweetening agent under the trade name Dulcin. It has the structure



It was therefore of interest to investigate the taste properties of the sulphur analog of Dulcin, that is, para ethoxy phenyl thio carbamide.

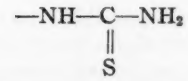


This compound was doubly interesting to investigate because some forty-six years ago Berlinerblau (*J. Prakt. Chem.* **2,** 30, 108) described this compound as being very bitter, and more recently Lange and Reed (*J. A. C. S.* **48,** 1070) stated that the replacement of oxygen by sulphur in many compounds caused bitterness, the most striking compound cited by them being Dulcin. Therefore this compound was prepared and it was found to exhibit this same peculiar property. A wide variety of other aryl thio carbamides were then prepared, of which the following are some of the more interesting.

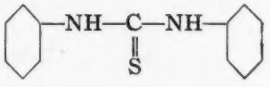


These were all shown to possess the same property.

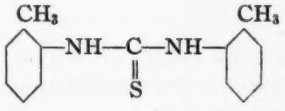
It was apparent, therefore, that the portion of the molecule rendering taste was the aliphatic portion, viz.,

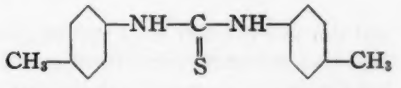


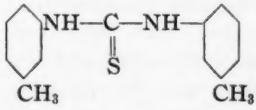
The question then presented itself as to what effect on the taste properties a change in the structure of this portion of the molecule would produce. The simplest method of changing this structure would be the replacement of one of the hydrogen atoms of the $-NH_2$ group with an aryl group. This would produce symmetrical diaryl thio carbamide, the simplest of which is symmetrical diphenyl thio carbamide (thio carbanilide). This compound was prepared by Hofmann (*Ann* **70,** 147 (1849)) some seventy-two years ago and described by him as having an intensely bitter taste. Lange and Reed (loc. cit.) also refer to this compound as being extremely bitter. This compound

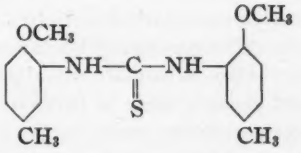


was prepared and the very surprising result noted that it, too, had this same peculiar taste difference according to the individual. It was found to be absolutely tasteless to certain people, while to others it had the intensely bitter taste described by Hofmann. Several other symmetrical diaryl thio carbamides of the following structures were examined,



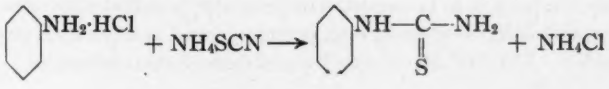




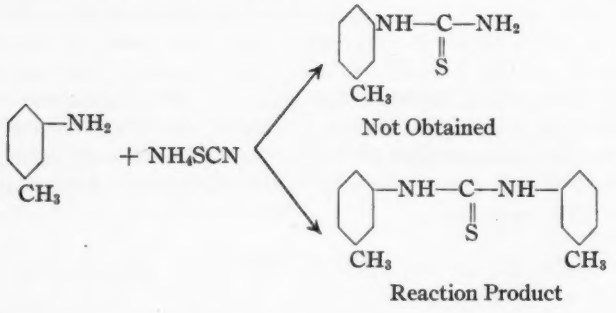


and all these were shown to possess this property.

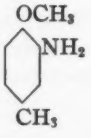
At this point a slight degression should be made to point out a very peculiar chemical reaction that occurred in the preparation of these compounds. The aryl thio carbamides were made from ammonium thio cyanate and a primary amine hydrochloride



It was shown, however, in two cases that under exactly similar conditions not a mono aryl thio carbamide but a diaryl thio carbamide was obtained. Thus when m-toluidine was heated with ammonium thio cyanate under usual conditions sym-di m-tolyl thio carbamide was obtained and not the expected m-tolyl thio carbamide.

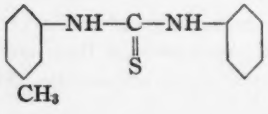


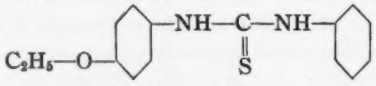

The other amine which gave this peculiar behavior was cresidine,



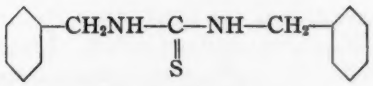

and this product also has a methyl group meta to the amine. It is probable that this meta substituent is responsible for this peculiar behavior and the author proposes to investigate this interesting point.

As symmetrical di-aryl thio carbamides were shown to possess this taste differentiation it became of interest to see whether unsymmetrical diaryl thio carbamides would also show this. In view of the wide structural changes made in previous compounds it was to be anticipated this taste peculiarity would persist and it was shown to do so. The following were made and examined, among others



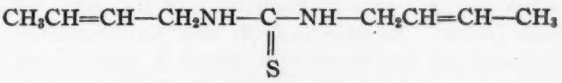




All the compounds so far reported on were aryl or diaryl thio carbamides. It became of interest therefore to examine some aralkyl and alkyl thio carbamides. The first investigated was di benzyl thio carbamide



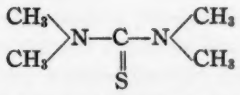
which was shown to have a similar property.

Di-crotyl thio carbamide

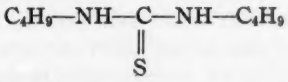


was shown to have a slightly bitter taste to non-tasters, but an extremely bitter taste to tasters.

Tetra methyl thio carbamide

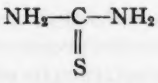


was shown to have a sour but not lasting taste to both tasters and non-tasters, and di-butyl thio carbamide



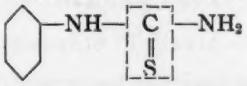
was somewhat bitter to both classes.

Thio carbamide itself



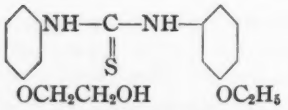
has a nauseating taste, not bitter but sour, to both tasters and non-tasters alike.

From the foregoing discussion it has become apparent that the taste of these compounds is closely connected with the C=S linkage, viz.,



because although the aryl group may be changed to many others this peculiar property remains. Also if a hydrogen of the $NH_2$ group is replaced with aryl the property remains. The fact that Dulcin, a sweet compound, is formed when the sulfur in *p*-ethoxy phenyl thio carbamide is replaced by oxygen indicates the taste-giving portion of the molecule is in the C=S linkage.

It is believed that this peculiar phenomenon is due to the difference in solubility of the products in the saliva of various individuals. For example, one product was examined




which was found to be tasteless to both tasters and non-tasters. This compound, however, is extremely insoluble, notwithstanding the hydroxy ethoxy group present. Another example was noted containing a solublizing group and this was found to be bitter to both tasters and non-tasters. Some experiments by A. F. Blakeslee have also indicated it is due to a difference in solubility. If this is the case, however, it remains just as inexplicable why these products should be insoluble in the saliva of some individuals and soluble in others. One possible explanation, offered with no experimental proof, is that the extremely small quantity necessary to produce taste is soluble in the saliva of all individuals, while the non-tasters have in their saliva a product, possibly a protein or colloid, which precipitates this as a very insoluble product and thus causes no taste to be sensed. One argument against this hypothesis is that non-tasters can place large quantities of these products on their tongues and not experience any taste. It would seem that the excess of reagent would precipitate all available protein and would then give a sense of taste, a sort of delayed taste. This, in fact, is experienced by some people, but whether this is due to the explanation offered is not known.

Inasmuch as this phenomenon is somewhat similar in its manifestations to color blindness it has been designated taste blindness. It is admitted that there is not an exact scientific analogy but the term taste blindness so effectively describes this phenomenon its use is believed permitted.

---

## *GENETICS OF SENSORY THRESHOLDS: TASTE FOR PHENYL THIO CARBAMIDE*

By Albert F. Blakeslee

Carnegie Institution of Washington, Department of Genetics, Cold Spring Harbor, N. Y.

Read before the Academy Tuesday, November 17, 1931

Dr. A. L. Fox[1] first showed that many people cannot detect the bitter taste in crystals of phenyl thio carbamide. In an earlier publication, Salmon and the writer[2] showed that inability to taste the crystals appears to be inherited as a Mendelian recessive. (The same conclusion has been reached independently by L. H. Snyder.[3]) In addition to "non-tasters" of the crystals, we were able to classify the "tasters" of the crystals roughly

according to their taste acuity by means of dilutions at which the bitter taste was first detected. This work on the genetics of taste thresholds has been extended, and the results on testing 103 families is presented in the present paper.

Table 1 summarizes the tests regarding "tasters" and "non-tasters." When both parents were non-tasters all the children have been non-tasters. When both parents were tasters, most of the children have been tasters but also a considerable percentage have been non-tasters. These relations would be expected if the difference between tasters and non-tasters were due to a pair of Mendelian factors. If all the tasters were heterozygous dominants one would expect a 3:1 ratio in the mating $T \times T$ and a 1:1 ratio in the back-cross $O \times T$. The deficiency in the non-taster class would be explainable by the frequent occurrence of parents homozygous for the factors for tasting. The critical mating for the hypothesis is that between two non-tasters ($O \times O$). If inability to taste the substance is due to a recessive factor, all the children should be non-tasters. Our tests from this mating furnish 22 children all of whom are non-tasters. Snyder's published results[3] add 17 more, making a total of 39. So far as the tests have gone one is warranted in making certain predictions regarding parents and offspring. A taster child will have at least one parent who is a taster and all the children of two non-taster parents will be non-tasters.

TABLE 1

HEREDITY OF TASTE DEFICIENCY FOR PHENYL THIO CARBAMIDE

(Percentages in parentheses)

| PARENTS | | CHILDREN | | |
|---|---|---|---|---|
| TYPE OF MATING | NO. OF MATINGS | NON-TASTERS ($O$) | TASTERS ($T$) | TOTAL |
| $O \times O$ | 10 | 22 | 0 | 22 |
| | | (100) | (0) | |
| $O \times T$ | 39 | 32 | 42 | 74 |
| | | (43.2) | (56.7) | |
| $T \times T$ | 54 | 22 | 109 | 131 |
| | | (16.8) | (83.2) | |
| Totals | 103 | 76 | 151 | 227 |
| | | (33.5) | (66.5) | |

In tables 1, 2 and 3, 6 persons were recorded as both parents and children.

The heredity differences between people in their taste perceptions of phenyl thio carbamide are not so simple as the preceding paragraph might lead one to suppose. Innate differences have been found to exist in regard to the weakest concentration (threshold) at which the substance could be detected, the apparent strength of the sensory reactions, the ability to detect differences in concentration and ability to distinguish

the taste of the thio compound from that of other substances such as acids for example.

Table 2 gives the thresholds for the families tested. The technique has been already described. Solutions were made up in artesian well water from a stock solution of 1:5000. The stock solution kept its strength without sensible deterioration for a considerable length of time. The solutions in the small bottles from which the tests were made were frequently renewed since they tended to lose their strength, perhaps on account of soluble substances on the soda fountain straws used in the test. About 0.6 cc. was usually taken up in the straw pipettes.

TABLE 2

PHENYL THIO CARBAMIDE

| TYPE OF MATINGS | PARENTS NO. OF MATINGS | TASTE THRESHOLDS OF 227 CHILDREN | | | | | |
|---|---|---|---|---|---|---|---|
| | | *O* | *A* (1:5,000) | *B* (1:20,000) | *C* (1:80,000) | *D* (1:320,000) | TOTALS |
| *O* × *O* | 10 | 22 | | | | | 22 |
| *O* × *A* | 2 | 4 | | | | | 4 |
| *O* × *B* | 8 | 9 | 1 | 4 | | | 14 |
| *O* × *C* | 25 | 18 | | 10 | 19 | 4 | 51 |
| *O* × *D* | 4 | 1 | | | 2 | 2 | 5 |
| *A* × *C* | 1 | | | 3 | 1 | | 4 |
| *B* × *B* | 5 | 2 | 1 | 3 | 2 | 2 | 10 |
| *B* × *C* | 12 | 8 | | | 10 | 9 | 27 |
| *B* × *D* | 2 | 2 | | | 4 | 4 | 10 |
| *C* × *C* | 17 | 5 | | | 19 | 16 | 40 |
| *C* × *D* | 12 | 3 | | | 10 | 12 | 25 |
| *D* × *D* | 5 | 2 | | 1 | 5 | 7 | 15 |
| Totals | 103 | 76 | 2 | 21 | 72 | 56 | 227 |

To avoid the psychological influence of expectation, it was a rule not to allow those about to be tested to see the test given to others. A bottle of plain water was of value at times when the subject was in doubt of his reaction or when psychological rather than taste reactions were suspected. In order to make the tests more closely comparable, all the recorded tests except two[5] were made by the writer. The first bottle contained a 1:1,280,000 solution in which concentration none was able to detect a bitter taste. Each solution tested was four times as strong as that previously tested. The threshold was recorded as the first concentration at which the taste was distinct.

Frequently the subject would say that a certain concentration was not water but could not decide what the taste was, even when the four tastes sweet, bitter, sour and salty were suggested to him. At the next higher concentration the subject might at once call the taste bitter and sometimes say that he now knew it was a bitter taste that he had been doubtful about in the more dilute solution. In such cases the threshold

was recorded for the stronger concentration at which the subject first recognized the taste was bitter. The subject was generally given also a grade stronger than his threshold if he did not object to the bitter taste. Those who had heard about the test and therefore expected a bitter taste may have scored a lower threshold than those who knew nothing about the chemical that was being given them. Tests of a few individuals at different times of day with a series of solutions in which each was twice as strong as the one previously tested indicated that environmental factors might have some slight influence upon the tasting ability. These differences were not great, however. Greater accuracy in determining thresholds and a smoother curve of distribution would have been possible if the factor 2, or 3 instead of 4, had been used in making up the solutions. This change, however, would have doubled the number of bottles needed for a test and, although it would have been better for a laboratory experiment, it might have been too inconvenient for use in survey work among non-scientific families where the technique employed must be simplified. A larger amount of fluid at each concentration would have had similar advantages and disadvantages.

Despite the shortcomings in method, some of which have been discussed, and the relatively small numbers for a classification of this kind, it is possible to draw some definite conclusions from the data assembled in table 2. All types of matings, for which there is more than a single family represented, have given some negative children. Homozygous cannot be separated from heterozygous tasters by their thresholds, nor by the use of di ortho tolyl thio carbamide, to which some "tasters" are insensitive. In matings with *O*'s, the effect of the grades of parents upon offspring is most clearly marked. $O \times O$ as well as $O \times A$ give only *O* offspring. $O \times B$ give about half as many tasters as *O*'s. $O \times C$ give about twice as many tasters as *O*'s and $O \times D$ give about four times as many tasters as negative children. The *O* grade forms the bottom of a series not only in respect to taste acuity of the individuals but also in respect to the effect upon the proportion of negative children as well as the proportion of acute tasters in the offspring. The threshold *A* appears to behave like a *O*, but if a considerable number, instead of only 4 children, had been recorded from this mating, some tasters might have appeared among the offspring. As will be seen later, there are probably different grades even of *O*'s and, considered from the standpoint of a continuous series from *D* to *O*, it would not be surprising that, if a very large number of $O \times O$ matings were studied, they might be found to yield a few taster offspring.

The acuteness of taste among the taster children also increases with increase of taste acuity in one of the parents. Thus from $O \times B$ matings, all the tasters are *B* or poorer tasters. From the mating $O \times C$, nearly a third of the tasters are *B*'s, two-thirds *C*, and about an eighth in the most

acute grade D. From the mating $O \times D$ there are no $B$ tasters and half of the tasters belong to the acute grade $D$.

When both parents are tasters, the effect of their thresholds upon the grades of their offspring is not so clear. The mating $B \times B$, however, is seen to be set apart from the matings of higher grade by having a large proportion of poor tasters. The matings of higher grade all have about the same proportion of acute (D) and medium (C) tasters in their offspring. The proportion of negative offspring, however, is somewhat different. If we disregard the single mating $A \times C$, and group together the matings $B \times B$, $B \times C$ and $B \times D$ we have 35 tasters to 12 negative children; a ratio of 3:1. Taking the matings of higher grade ($C \times C$, $C \times D$ and $D \times D$) we have in the offspring 70 tasters to 10 non-tasters; a ratio of 7:1 or half as many non-tasters as from the group of lower grade parents.

In table 3 is given a summary of the threshold tests of table 2 classified under parents and children as well as according to male and female. The data suggest that more of the children than of the parents are acute tasters (grade D). This is true for both males and females. There are factors which might tend to make children score lower than their parents for the same taste acuity. Young children are not as familiar as their parents with bitter tastes. As we have pointed out earlier, familiarity with a taste will probably lead to a lower recorded threshold for the same sensation. To avoid as much as possible the difficulties of language it has been the rule not to include in the records children under 10 years of age. Older children often have difficulty in describing the taste but when asked to compare it with something they have tasted before, country children are likely to say it tastes like dandelion stems which they have stuck in their mouths in making "curls" while city children are likely to compare the taste with some medicine they have taken. The data in table 3 also suggest that there are more acute tasters among females and this difference appears in both parents and offspring. The proportions of non-tasters among males and females are reversed in parents and children and it is questionable how much significance should be attributed to apparent differences between parents and offspring and between males and females in respect to taste acuity. Age is not necessarily associated with poor taste since a subject 82 years old could detect bitterness in a 1:320,000 dilution.

The grand totals at the bottom of table 3 indicate that there is a curve of distribution with a discontinuity between the "non-tasters" (O) and those with a threshold at 1:5000. A further word is desirable at this point in regard to the manner of making the tests. A preliminary test had shown that there were few (12%) who had a threshold as high as 1:20,000. The most concentrated solution, therefore, was taken at 4

times as strong, or 1:5000. Later a few persons (1%) were found who could not taste the bitterness until the 1:5000 solution. Our practice was to try the crystals if the subject could not taste the 1:5000 solution. The fact that a few could taste bitterness in the solutions even in high (D) dilutions but were unable to taste the crystals suggested trying saturated solutions of the compound on negative subjects. The majority of non-tasters thus retested could detect bitterness in a cold saturated solution while the rest could detect it in a hot saturated solution with crystals in suspension or better in a similar hot solution in weak alcohol. Dr. Fox has found phenyl thio carbamide to be 0.26% soluble in water at

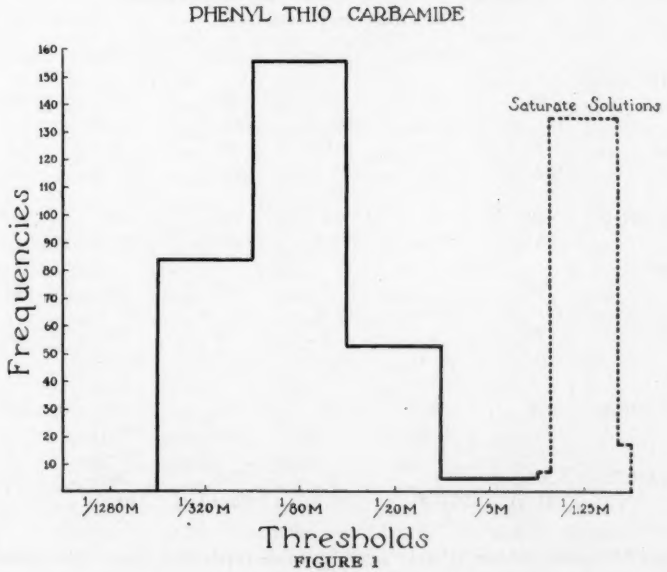



FIGURE 1

18°C. and 5.93% at 100°. In 95% alcohol it is 5.6% soluble at 16°C. and 68% in boiling alcohol. The fact that all who could not taste the crystals whom we have retested with saturated solutions have been found to be positive (25 cases) indicates that all persons can taste bitterness in the compound if only it can be gotten to the taste organs in a sufficiently concentrated condition. In a diagram (Fig. 1) we have indicated the "non-tasters," as previously defined, by a dotted line. Taste ability for the substance therefore forms a bimodal curve. The outer hump on the dotted line represents those who needed a hot saturated solution to sense the bitter taste. The inner hump represents a subject who tasted bitterness in the crystals when a large amount was taken and who had a thresh-

old of 1:5000 after chewing gum for a couple of hours. Gum chewing did not change taste of other *O*'s, however. Some *O*'s had no bitter taste from an excess of the dry crystals while some got a feeble taste under these conditions.

The fact that in the case of phenyl thio carbamide an apparent deficiency in taste has been resolved into merely a difference in thresholds suggests that threshold differences may be responsible for other apparent deficiencies in taste and smell, such as the differences in peoples' reactions to odors of Verbenas earlier discovered by the writer.[4]

TABLE 3

TASTE ACUITY FOR PHENYL THIO CARBAMIDE

(Percentages in parentheses)

| | *O* | *A* 1:5000 | *B* 1:20,000 | *C* 1:80,000 | *D* 1:320,000 | TOTALS |
|---|---|---|---|---|---|---|
| PARENTS | | | | | | |
| Male | 26 | 3 | 16 | 48 | 10 | 103 |
| | (25.2) | (2.9) | (15.5) | (46.6) | ( 9.7) | |
| Female | 33 | 0 | 16 | 36 | 18 | 103 |
| | (32.0) | (0) | (15.5) | (35.0) | (17.5) | |
| Totals, parents | 59 | 3 | 32 | 84 | 28 | 206 |
| | (28.6) | (1.5) | (15.5) | (40.8) | (13.6) | |
| CHILDREN | | | | | | |
| Male | 44 | 1 | 9 | 34 | 19 | 107 |
| | (41.1) | (0.9) | (8.4) | (31.8) | (17.8) | |
| Female | 32 | 1 | 12 | 38 | 37 | 120 |
| | (26.7) | (0.8) | (10.0) | (31.7) | (30.8) | |
| Totals, children | 76 | 2 | 21 | 72 | 56 | 227 |
| | (33.5) | (0.9) | (9.3) | (31.7) | (24.7) | |
| Grand totals | 135 | 5 | 53 | 156 | 84 | 433 |
| | (31.2) | (1.2) | (12.2) | (36.0) | (19.4) | |

The *O*'s in the tables, it may now be said, represent those who have a threshold above 1:5000. The results would be about the same if the *O*'s were defined as those who cannot taste the crystals.[5] The factors responsible for the bimodality of the curve in figure 1 and for the fact that some who have low thresholds for bitterness in the solutions cannot taste the crystals are not entirely clear. Probably solubility of the compound in the saliva may have something to do with the phenomena. From the pH of salivas shown in table 4, it is obvious that acidity of saliva is not the controlling factor. These pH determinations were kindly made by Miss Satina by the use of brom thymol blue as an indicator. These pH's probably represent innate differences as well as possible effects of environmental factors.

The inquiry has frequently been made whether or not acuteness of taste

for the thio compound indicates acuteness of taste for other substances. This was early seen not to be the case since those in a family who were said to be the first to detect sourness in milk were sometimes good and sometimes poor tasters for the thio compound. It seemed desirable, however, to compare acuity of taste for different bitter compounds as well as for representatives of the other three primary tastes, sweet, sour and saline. Accordingly, the tests shown in table 4 were made of 21 individuals all of whom except No. 17 were connected in some way with the Department

TABLE 4

THRESHOLDS IN TASTE ACUITY

| SUBJECT NO. | BITTER: P. THIO CARB. | BITTER: QUININE | BITTER: PICRIC ACID | SOUR HCl (38%) | SWEET SACCHARINE | SALT NaCl | ACIDITY OF SALIVA (pH) |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 1:5M | 1:20M | 1:800 | 1:20M | 1:400 | 7.0 |
| 2 | 0 | 1:5M | 1:20M | 1:1600[a] | 1:80M | 1:400 | 6.8 |
| 3 | 0 | 1:20M | 1:20M | 1:200 | 1:20M | 1:400 | 6.6 |
| 4 | 0 | 1:20M | 1:20M | 1:400 | 1:10M | 1:400 | 7.1 |
| 5 | 0 | 1:20M | 1:20M | 1:1600[a] | 1:20M | 1:400 | 7.0 |
| 6 | 0 | 1:20M | 1:80M | 1:800 | 1:40M | 1:400 | 6.8 |
| 7 | 0 | 1:20M | 1:80M | 1:800 | 1:40M | 1:400 | 7.0 |
| 8 | 0 | 1:80M | 1:80M | 1:800 | 1:40M | 1:400 | 6.8 |
| 9 | 0 | 1:80M | 1:80M | 1:1600 | 1:40M | 1:400 | 6.2 |
| 10 | 1:20M | 1:80M | 1:20M | 1:1600 | 1:40M | 1:400 | 6.9 |
| 11 | 1:80M | 1:5M | 1:20M | 1:400[b] | 1:10M | 1:200[c] | ... |
| 12 | 1:80M | 1:5M | 1:320M | 1:400 | 1:40M | 1:200[c] | 6.9 |
| 13 | 1:80M | 1:20M | 1:80M | 1:1600 | 1:40M | 1:200 | 7.0 |
| 14 | 1:80M | 1:20M | 1:80M | 1:1600 | 1:80M | 1:400 | 6.7 |
| 15 | 1:80M | 1:20M | 1:320M | 1:400 | 1:80M | 1:200 | 6.9 |
| 16 | 1:80M | 1:320M | 1:80M | 1:1600 | 1:80M | 1:800[c] | 7.1 |
| 17 | 1:80M[d] | 1:320M[d] | 1:80M | 1:1600[d] | 1:80M | ... | ... |
| 18 | 1:80M | 1:320M | 1:320M | 1:1600 | 1:40M | 1:400 | 6.9 |
| 19 | 1:80M | 1:320M | 1:320M | 1:1600 | 1:80M | 1:400 | 6.5 |
| 20 | 1:320M | 1:20M | 1:80M | 1:800 | 1:40M | 1:400 | 6.8 |
| 21 | 1:320M | 1:20M | 1:80M | 1:1600[b] | 1:40M | 1:400 | 6.7 |
| Dilution factors | 4 | 4 | 4 | 2 | 2 | 2 | |

[a] All solutions of HCl taste like alum, not sour. [b] This solution of HCl tastes astringent, stronger solutions taste sour. [c] 1:200 and 1:100 NaCl taste sour and astringent. [d] Solutions of quinine, phenyl thio carbamide and HCl taste alike. [e] This solution of NaCl tastes sour, next stronger solution tastes salty.

of Genetics. The tests were made with soda fountain straws and efforts were made to avoid influence of previous tastes by taking the tests of different substances at different times or by drinking water between tests when a longer separation in time was not convenient. The method employed is somewhat crude and subject to errors already discussed in tests of the thio compound. The table gives information, however, that

appears not to have been assembled elsewhere. No one of the group is the poorest taster in respect to all the substances tested and no one is in the best grades for all the substances. Some, however, are relatively acute tasters for all substances tested and others are relatively poor for all. It was surprising further to learn that there was no close connection between ability to taste two different bitter substances. Some of these negative for the thio compound have relatively low thresholds (1:80,000) for quinine sulphate and for picric acid. Tests on a larger number might show that some who were negative for the thio compound were in the most acute grade for quinine and picric acid. The most striking difference in reaction to a bitter substance is shown by subject No. 12, who was in the poorest grade for quinine but in the most acute grade for picric acid. It had earlier been seen by the use of powdered quinine sulphate, as an attempted standard for the word bitter, that the reaction to quinine did not run parallel to that for the thio compound. Incidentally two cases were found who were apparently negative to powdered quinine sulphate. It would be interesting to test out the taste buds for different bitters.

For sour, sweet and salt test substances it was necessary to use a dilution of 2 instead of 4 to get a reasonable spread to the thresholds. Salt showed the least differences between the reactions (the highest threshold was only 4 times the concentration of the lowest) despite the differences between people in respect to the amount of salt they like in their food. The threshold at which a substance is detected does not of itself give information regarding likes or dislikes of the subject nor does it indicate the strength of sensation felt. Some were able to detect the thio compound even in the lowest concentration (1:320,000) but found little difference in sensation between this and the 1:5000 solution which is 64 times stronger. A few could not detect the compound until they were given the relatively strong solution of 1:20,000 but at this threshold made an extremely strong complaint of the bitter taste. There appears to be no close correlation between thresholds and emotional response.

The saccharine had a fleeting sweet taste at the thresholds and was difficult to record. The highest threshold was 8 times the strength of the lowest. The dry powder tastes both sweet and bitter to most subjects. For the sour hydrochloric acid the lowest threshold was 8 times as dilute as the highest. The sour sensation was easier to grade than that from saccharine and salt. Despite the recognized differences between members of a household in their ability to detect sourness in milk and also similar differences in professional milk testers, several dairy investigators who were consulted believed that no comparable tests of milk tasters had ever been made to discover how they differed in taste perceptions. The bitter substances show greater differences between the high and low thresholds than the other substances tested.

It is probable that the differences shown in table 4 represent innate heritable differences in the subjects tested, although this has been established only for the phenyl thio carbamide. By tabulating the thresholds of parents and offspring for other substances than the thio compound, it may be possible to demonstrate an hereditary basis for other taste perceptions. Certain weakly acid salts have been found tasteless to a person with a high threshold for HCl.

From the notes to table 4, it will be seen that there were some qualitative differences in reactions to the same substance, which need further study. In earlier tests, it was not infrequent to have the taste of the thio compound described by some other term than bitter, such as sour, sweet, astringent, like lemons, rhubarb, cranberries or vinegar. They have been recorded as tasters under the appropriate threshold. It has been possible as yet to retest carefully only a few of such cases. Some have been shown to be due to using the wrong term for the taste since these subjects could distinguish bitter from sour substances when they had these substances given them in comparable tests. There were several cases found, however, which were clearly due to what may be called *bitter-sour indiscrimination.* In the first (no. 17 of table 4) quinine sulphate, phenyl thio carbamide and hydrochloric acid were all described as puckery or astringent plus a taste like vinegar. Picric acid was said to have the same taste without the astringency. It will be noted that subject No. 17 detected the acid and the bitters in low concentrations although (except for the picric acid) he could not discriminate what most call bitter from what most call sour. A second subject described hydrochloric acid and the thio compound as lemony in taste but the quinine and picric acid as sour. A third, an assistant in a western department of zoölogy, was retested by a fellow assistant. She reports that quinine and hydrochloric acid are indistinguishable to her. When the solutions are strong she calls them both bitter, when they are dilute she calls them sour. Sour and bitter are terms to denote to her quantitative and not qualitative differences. Perhaps a majority of the bitter-sour indiscriminators would react like this subject and call the thio compound bitter rather than sour. It is probable therefore that a considerable number among those calling the substance bitter would be shown by tests to be examples of taste indiscrimination. There is also evidence of various grades of acuteness in taste discrimination.

A further investigation of taste indiscriminations, which appear to correspond to color blindness in vision, is deferred to a later study. Such indiscriminations in taste appear not to have been noted before. Parker in his interesting book on taste and smell[6] makes no mention of them. A number of physiologists and psychologists who have been consulted are also unaware of mention of such phenomena in the literature. That

they have not been discovered before may be due to the fact that the emphasis has generally been placed upon the reactions of a single individual rather than upon the sensory differences between different individuals.

Differences in taste thresholds for a number of other substances than phenyl thio carbamide have been found by us and other investigators.[7] The same is true for odors. In the single case investigated we have found these differences in powers of sensory perceptions innate and hereditary. Evidence is thus given for the belief that humans are born with innate differences in respect to all their senses and that different people live in different worlds, therefore, so far as their sensory reactions are concerned.

[1] Fox, A. L., immediately preceding paper in this issue, These PROCEEDINGS. Cf. also note in *Science News Letter*, April 18, 1931.

[2] Blakeslee, A. F., and Salmon, M. R., *Eugenical News*, **16**, 105–108, (July, 1931).

[3] Snyder, L. H., *Science, N. S.*, **74**, 151–152, August 7 (1931). Dr. Snyder used para-ethoxy-phenyl-thio-carbamide, a slightly different substance from that which we used.

[4] Blakeslee, A. F., *Science, N. S.*, **48**, 298–299, September 20 (1918).

[5] One child from the mating $C \times C$ who could not taste the 1:5000 solution had a delayed weakly bitter taste from the crystals and 2 children from the mating $O \times C$ showed a similar reaction. One parent in the $O \times O$ mating, who at first reported both solutions and crystals to be tasteless, later, after others in his family had been tested and he had heard their discussions, said the crystals tasted bitter. Other than these mentioned none could detect bitter in the crystals who did not taste bitter in the 1:5000 solution. Retests with a large amount of the crystals have given a weak bitter taste to some recorded as negative. All the parents in table 2 were tested with solutions as well as all their children except 4 children from the mating $O \times O$ who were tested with crystals. Two of the latter were tested by some one else. Otherwise all the records in the table were obtained by the writer personally. To the $O \times O$ mating might have been added 4 negative children tested with both solutions and crystals, the parents of whom were tested with crystals by the writer's assistant. The records in the tables are not entirely comparable with those in our previous paper[2] in which the $O$'s include those who tasted the solutions but found the crystals tasteless. A few errors in records in the earlier tables have been corrected.

[6] Parker, G. H., *Smell, Taste and Allied Senses in Vertebrates*, Lippincott & Co., 1922.

[7] Biester, Alice, Wood, Mildred W., and Wahlin, Cecile S., on sugars, *Am. J. Physiol.*, **73**, 387–396 (1925). Ward, J. C., and Munch, J. C., on strychnine, *J. Am. Pharm. Assoc.*, **19**, 1057–1060 (1930). Munch, J. C., on capsicums, *J. Am. Pharm. Assoc.*, **18**, 1236–1246 (1929).

# ELIMINATION OF SELF-STERILITY IN THE STYELA EGG—A RE-INTERPRETATION WITH FURTHER EXPERIMENTS

BY HAROLD H. PLOUGH

DEPARTMENT OF BIOLOGY, AMHERST COLLEGE

Communicated December 1, 1931

In a previous communication[1] I noted that the Ascidian *Styela partita* may exhibit complete or partial self-sterility, and experiments were cited indicating that this self-sterility was eliminated by adding to the sea water in which the eggs were fertilized small amounts of $NH_4OH$ or NaOH. Since only normally shed eggs can be used in this species and since sperm are passed into the sea water simultaneously, all the eggs used have already been exposed to sperm from the same animal before any experimental crosses can be made. After sperm suspensions were taken, the dishes were stirred to secure an even exposure to "own" sperm, and equal samples of eggs placed in 10 c. cm. of sea water. To these were added equal amounts of the various sperm suspensions, including the own sperm. The results of crossing were thus doubly controlled—by a sample of eggs from the original dish as well as by one with an added amount of own sperm.

Styela aggs often give no evidence of self-sterility, a high percentage developing from the start with no increase on crossing. When, on the other hand, they show no development with own sperm—which is found to be effective in crosses— but positive results with sperm of other individuals, the animal producing them may be called completely self-sterile. The more usual result is that a certain number of eggs develop following self-fertilization, but much higher percentages of development following exposure to sperm suspensions from other individuals. Such animals are said to be partially self-sterile. We may speak of the elimination of self-sterility in either case. When eggs which have shown complete or partial self-sterility subsequently give as great a percentage of successful fertilization with own sperm as do any of the crosses, the *self*-sterility has been eliminated. It is clear that complete *self*-fertility and complete (100%) *fertility* may be quite different things.

During the past summer at Woods Hole I have made a large number of attempts to eliminate self-sterility by the use of alkaline sea water, but, contrary to last year's work, with generally negative results. Duplicate sets of egg samples were made, and selfed or crossed with the sperm suspensions available. Immediately afterward 4 drops of a N/10 solution of $NH_4OH$ or NaOH were added to each dish of one set. The results were variable but showed most frequently a slightly *reduced* percentage of fertilization as compared with the controls in both selfed and crossed

samples. The technique was varied without significant differences in the sea water after varying intervals the only change noted was a greater number of normal larvae and not a larger total number of developing eggs. Systematic tests of fertilization in sea water ranging from pH 7.5 to pH 8.5, using HCl and NaOH, showed with increasing clearness that the normal pH of sea water—approximately 8.2—gave maximum percentages of fertilization, and the greater the departure from normal in either direction the lower the percentages. The differential between self and cross-fertilization—that is, the degree of self-sterility—remained approximately the same in the majority of tests. Obviously my conclusion based on the experiments of last year—that self-sterility can be eliminated by fertilizing in alkaline sea water—was wrong.

Yet last year's results were secured in a considerable number of experiments, and while less extensive than those of the present year since the treatment was tried late in the breeding season, nevertheless were per-

TABLE 1

| | 8.00 | 8.30 | 9.00 | 9.30 | 10.00 | 10.30 | Time |
|---|---|---|---|---|---|---|---|
| *Ee* | 0.24 | 0.12 | 0.37 | 0.53 | 0.47 | 0.56 | |
| $Ee_1$ | 0.16 | 0.15 | 0.33 | 0.30 | 0.32 | 0.45 | |
| *Ed*, | 0.48 | 0.41 | 0.40 | 0.43 | 0.51 | 0.54 | |
| $Ed_1$ | 0.27 | 0.17 | 0.37 | 0.50 | 0.47 | 0.39 | |

Eggs and sperm suspensions

TABLE 2

| | 9.50 | 10.50 | 11.50 | Time |
|---|---|---|---|---|
| *Ff* | 0.07 | 0.11 | 0.24 | |
| *Fa* | 0.72 | 0.69 | 0.41 | |

Eggs and sperm suspensions

fectly definite and uniform. The problem thus presented itself: since a change in the hydrogen-ion concentration of the sea water is ineffective in eliminating self-sterility, how were the positive results of last year's experiments brought about. The answer to this question has brought to light the error usually inherent in a faulty experiment. In the experiments of last year there were two variables, only one of which was controlled. The loss of self-sterility turns out to be determined by the second variable, while the one actually assigned as the cause is negative in its action. A comparison of the methods used in the two series of tests gave the key to the solution of the problem. Last year my practice was to prepare one set of egg samples, fertilize, and set these aside as the controls. A second set was then prepared with the alkaline solution added. The eggs used in the experimental series had thus stood in the dishes longer than the controls, and when a large series of crosses was made this interval was considerable.

In the experiment tabulated as table 1 in my earlier paper, my notes show that 45 minutes elapsed between the fertilization of the first set of egg samples (K), and that of the second ($K_1$) in which self-sterility was found to have disappeared. In this year's tests the egg samples were prepared at the same time and fertilized almost simultaneously. It is this time interval which is the determining factor in self-sterility, and not the alkalinity of the solution. Eggs and sperm on standing in sea water gradually undergo some spontaneous change as a result of which the block to fertilization which was present at the time of shedding has disappeared.

The evidence for this solution of the problem was secured by making up

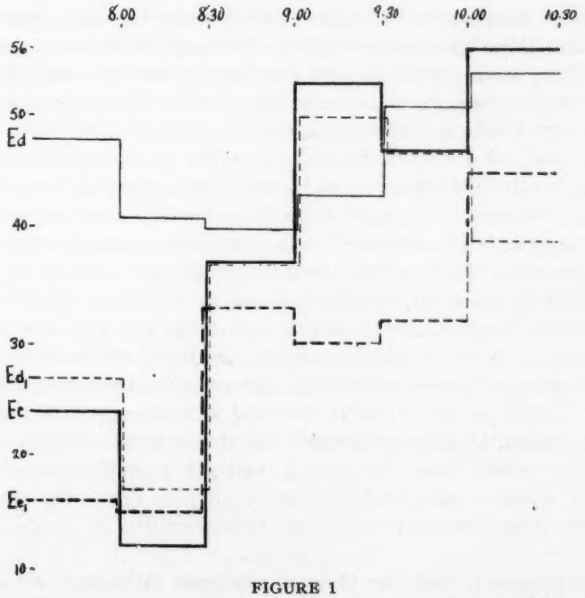



FIGURE 1

successive self and cross fertilization tests at intervals of one half an hour beginning as soon as possible after shedding. At each interval also a duplicate set of eggs was fertilized with 2 drops of N/10 NaOH added to each 10 c. cm. of sea water. Table 1 gives the results of one such experiment, and these are shown in the form of a graph in figure 1. The heavy line indicates the percentages of successful fertilization after selfing (*Ee*), while the light line gives the results of crossing (*Ed*). *E* eggs were partially self-sterile at first, but this self-sterility gradually disappeared after $1^1/_2$ to 2 hours. The dotted lines show in each case ($Ee_1$, $Ed_1$) the result when the final fertilizations were made in alkaline sea water. It is clear that in

general this treatment slightly decreases the percentage of fertilization as compared with the untreated lots. It is clear, also, that if the level of the dotted line *Ee*, after 1 hour had elapsed were compared with that of the corresponding full line *Ee* at the beginning, it would appear that NaOH had brought about a definite increase. Because of my failure to make complete tests at definite intervals after shedding, just this mistake was made in interpretating my data last year. In the light of the evidence given in figure 1, which is one of many tests showing similar results, it seems clear that the increase in fertility in the $K_1$ series (table 1 in my 1930 report) was due to the gradual loss of self-sterility on standing and not to the effect of the alkalinity of the sea water at all.

Certain of my experiments suggest that cross-fertility may increase during the first three hours after shedding, although much less rapidly than self-fertility. Such results are not conclusive, however, since it will be recalled that the eggs are standing in a suspension of own sperm up to the time the cross-fertilizations are made, and the increase in self-fertility alone might account for the upward trend. In order to secure additional data on this point all of the eggs of several animals were removed from the dishes into which they were shed, squirted into fingerbowls of fresh sea water, and allowed to settle. This treatment brings about very considerable dilution of the suspension of own sperm surrounding the eggs, and a truer picture of the effect of staling on cross-fertility may be expected. There is much individual variation in the results of such tests, but a typical series of percentages is given in table 2. *F* eggs gradually increased in fertility during 2 hours with own sperm (*Ff*), but cross-fertility gradually diminished. It is not possible to prove the point with the eggs of Styela, which must be exposed to own sperm before the staling process begins. Nevertheless the trends shown in table 2, and other similar tests, indicate that 3 or 4 hours after shedding the fertility of eggs with any sperm suspension is approximately equal, and that thereafter it rapidly falls off to 0.

It is of interest to find that these results agree with those of Fuchs[2] on the eggs of Ciona at Naples. In this species the eggs and sperm are taken directly from the genital ducts so that there is no initial exposure to own sperm. Fuchs found that self-fertility gradually increased after the eggs were transferred to sea water, and that cross-fertility diminished. The chief difference is that in Ciona cross-fertility apparently exceeds self-fertility for a much longer period with sperm suspensions of approximately equal concentration, while in Styela self-cross and fertility tend to equalize after 2 to 3 hours.

It is impossible to say whether the gradual loss of self-sterility in Styela is due exclusively to a change in the eggs or in the sperm. Self-sterility is a reciprocal incompatibility between gametes from the same individual.

The disappearance of it on standing seems to suggest a diffusion of some substance into the sea water from either or both types of gametes. It seems simplest to think of it as diffusing out from the eggs, after which all sperm are equally likely to enter, but proof of this is impossible to get with Styela where fertile eggs cannot be secured free of exposure to own sperm. Morgan's suggestion[3] that some secretion of the test cells immobilizes the sperm from the same animal also remains tenable, provided the test cells gradually lose their immobilizing secretion into the sea water.

Of one point it is possible to be certain, however. The spermatozoa which first reach the egg—whether they penetrate the membrane or not—are put out of action. They can be seen in large numbers around the membranes in the first selfed samples, and they remain there. Yet only a small number of these eggs develop. Later samples with the same original sperm surrounding the eggs shown a greatly increased percentage of development as a result of re-fertilization from the same sperm suspension. Obviously the later sperm arrivals penetrate the eggs and initiate development without hindrance. Just[4] has suggested that self-sterility may be due to the inhibitory action of blood, as shown by Lillie[5] and Just[6] in Echinoderm eggs. The data given in this report appear at first sight to justify this suggestion, since blood also diffuses out from the eggs, and they gradually increase in fertility. The parallelism falls down completely, however, in accounting for the specificity of self-sterility. In Echinoderms blood inhibits all sperms, but its effect is most extreme in out- (species) crosses. Self-sterile Ascidians shed gametes which are perfectly capable of development in out-crosses, but inhibition occurs when gametes from the same individual are mixed.

I am glad to acknowledge the painstaking assistance in this work of Miss Evangeline Alderman of Wellesley College.

[1] Plough, *Proc. Nat. Acad. Sci.*, **16**, 800 (1930).
[2] Fuchs, *Arch. Entw-mech.*, **40**, 205 (1914).
[3] Morgan, *Proc. Nat. Acad. Sci.*, **9**, 170 (1923).
[4] Just, *Protoplasma*, **10**, 300 (1930).
[5] Lillie, *J. Exp. Zoöl.*, **16**, (1914).
[6] Just, *Biol. Bull.*, **44**, 10 (1923).

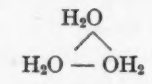
# THE PHYSIOLOGICAL EFFECT OF TRIHYDROL IN WATER

By T. Cunliffe Barnes

Osborn Zoölogical Laboratory, Yale University

Communicated December 14, 1931

It is well known that the molecules of water exist in various polymers. Water as a liquid is a system of three components (H. T. Barnes, 1928); ice, or trihydrol, which is present in greatest concentration at the freezing point; dihydrol, the main component at ordinary temperatures; and monohydrol, or steam, increasing as the temperature rises to the boiling point. The structural formula for trihydrol may be written



The paper of W. H. Barnes (1929) may be consulted for recent results on the crystal structure of ice as revealed by x-ray analysis. At any temperature there will be a certain relative proportion of all three of these substances. Even in water vapor there are still some polymerized molecules present. Living matter as it exists today is essentially an aqueous system and it is obvious that the properties of water as a liquid are of peculiar interest to the biologist. The writer has found that water containing relatively large quantities of the colloidal form trihydrol has a profound physiological effect.

The growth and photosynthesis of *Spirogyra* is greatly accelerated and the multiplication of Protozoa is more rapid in cultures rich in trihydrol than in controls of monohydrol and dihydrol at the same temperature. The writer has undertaken an extensive series of experiments on the trihydrol effect on a wide range of biological phenomena. Of these the effect on *Spirogyra* may be mentioned at this time. In series A the cultures contain no salts; in B Moore's solution is present in several concentrations and in C trihydrol and monohydrol water are dripping at a constant rate into the medium. Detailed tables of data will be published elsewhere but the results in series A are selected for the present report.

The data based on three of the cultures in series A are presented in table 1. These observations are supported by 25 duplicate experiments. In the case illustrated 500-cc. Pyrex beakers thoroughly cleaned were used; the tops were plugged with cotton; a constant temperature of 10° was maintained and the diffuse North light from a window was carefully measured in each culture with an illuminometer. The monohydrol, dihydrol and trihydrol water was renewed each day. The identical molecules may be changed at will into the various polymers (H. T. Barnes,

1928) and the effects duplicated, i.e., the same sample of water after giving the monohydrol effect was converted into trihydrol by a freezing process and the biological effect produced. After four days the standard colony of cells had grown to a volume of 14 cc. in the monohydrol and 132 cc. in the trihydrol. Extensive tables of data will appear elsewhere.

The effect on Protozoa will be the subject of a future report for the trihydrol influence on holozoic or saprozoic organisms is a special problem. *Spirogyra* on the other hand utilizes the water molecules in photosynthesis and the working hypothesis is put forward that the colloidal hydrol aggregates are the first building stones in the formation of carbohydrates.

TABLE 1

THE EFFECT OF WATER POLYMERS ON *Spirogyra*

| EXPERIMENT | TEMPERATURE | LIGHT INTENSITY | PREDOMINANT POLYMER | DIMENSIONS OF COLONY DEC. 3 | DIMENSIONS OF COLONY DEC. 7 |
|---|---|---|---|---|---|
| A1 | 10°C. | 30 ft. candles | Monohydrol | $4 \times 3 \times 1/2$ cm. | $5 \times 4 \times 3/4$ cm. |
| A2 | 10°C. | 30 ft. candles | Dihydrol | $4 \times 3 \times 1/2$ cm. | $6 \times 4 \times 1\,3/4$ cm. |
| A3 | 10°C. | 30 ft. candles | Trihydrol | $4 \times 3 \times 1/2$ cm. | $7 \times 7 \times 2\,3/4$ cm. |

The effect on other biological processes appears to be a catalysis phenomenon produced by the colloidal water particles. This effect will be discussed in a future publication.

There are many biological phenomena which may ultimately receive their explanation on the basis of the trihydrol effect. One may mention the anomalous increase in the plankton following the melting of the ice, the unexplained richness of micro-organisms in Arctic waters and the paleontological consequences of glacial epochs.

Barnes, H. T., *Ice Engineering*, Montreal, 1928, 500 pp.
Barnes, W. H., *Proc. Roy. Soc.* **A, 125,** 670–693 (1929).